Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/002325

International filing date: 24 January 2005 (24.01.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/538,682
Filing date: 23 January 2004 (23.01.2004)

Date of receipt at the International Bureau: 31 March 2005 (31.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

1299567

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 22, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/538,682*
FILING DATE: *January 23, 2004*
RELATED PCT APPLICATION NUMBER: *PCT/US05/02325*

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

EXPRESS MAIL LABEL NO.: *EL975241279US*

INVENTOR(s)/APPLICANT(s)

Given Name (first and middle [if any])	Family Name or Surname	RESIDENCE (City and either State or Foreign Country)
Fred R.	Hirsch	
Paul A.	Bunn, Jr.	
Wilbur A.	Franklin	
Chris	Coldren	
Razvan	Lapadat	
Michio	Sugita	
Mark	Geraci	
Barbara	Helfrich	
Samir	Witta	

[ ] *Additional inventors are being named on the ___ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

"GEFITINIB SENSITIVITY-RELATED GENE EXPRESSION AND PRODUCTS AND METHODS RELATED THERETO"

CORRESPONDENCE ADDRESS

Direct all correspondence to:

[X] Customer Number **22442** → Place Customer Number Bar Code Label here

OR *Type Customer Number here*

[X] Firm or Individual Name	Sheridan Ross P.C.				
Address	1560 Broadway, Suite 1200				
Address					
City	Denver	State	Colorado	ZIP	80202-5141
Country	United States	Telephone	(303) 863-9700	Fax	(303) 863-0223

ENCLOSED APPLICATION PARTS (check all that apply)

X	Specification	Number of Pages	50		CD(s), Number	
	Drawing(s)	Number of Sheets		X	Other (specify)	Postcard receipt.
	Application Data Sheet. See 37 CFR 1.76					

METHOD OF PAYMENT FOR FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

X	Applicant claims small entity status. See 37 CFR 1.27	FILING FEE AMOUNT ($)
	A check or money order is enclosed to cover the filing fees	$
	The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 19-1970	
	Payment by credit card. Form PTO-2038 is attached	

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: __

Respectfully submitted,

SIGNATURE: Angela Sebor

TYPED OR PRINTED NAME: ANGELA DALLAS SEBOR

TELEPHONE: (303) 863-9700

Date:	January 23, 2004
Registration No.	42,460
Docket No.:	2848-65PROV

Gefitinib Sensitivity-Related Gene Expression and Products and Methods Related Thereto

Background of the Invention

Lung Cancer is the leading cause of death from cancer worldwide. Chemotherapy is the mainstay of treatment for lung cancer. However, less than a third of patients with advanced stages of non-small cell lung cancer (NSCLC) respond to the best two chemotherapy drug combinations. Therefore, novel agents that target cancer specific biological pathways are needed.

The epidermal growth factor receptor (EGFR) is one of the most appealing targets for novel therapies for cancer. EGFR plays a major role in transmitting stimuli that lead to proliferation, growth and survival of various cancer types, including, but not limited to, NSCLC. Ligand binding to the EGFR receptor leads to homo- or heterodimerization of EGFR with other ErbB receptors. EGFR is overexpressed in a large proportion of invasive NSCLC and in premalignant bronchial lesions. Bronchioloalveolar carcinoma (BAC), a subtype of non-small cell lung cancer, represents the major form of lung cancer in non-smoking females and is rising in frequency, and epidermal growth factor receptor (EGFR) is expressed with high frequency in BAC. Unfortunately, the response of BACs to conventional chemotherapy is poor. Activation of the EGFR receptor leads to simultaneous activation of several signaling cascades including the MAPK pathway, the protein kinase C (PKC) pathway and the PI(3)K-activated AKT pathway. EGFR signaling translated in the nucleus leads to cancer cell proliferation and survival.

Targeted therapy against the EGFR receptor has produced response rates of 25-30% as first line treatment and 11-20% in 2^{nd} and 3^{rd} line settings (e.g., chemo-refractory advanced stage NSCLC). For example, in phase II clinical trials, 11-20% of patients with chemo-refractory advanced stage NSCLC responded to treatment with the EGFR tyrosine kinase inhibitor gefitinib (commercially available as Iressa®, ZD1839). A trial evaluating the activity of the EGFR inhibitor, erlotinib (Tarceva®, OSI-774) has been completed and the results will be reported in the near future. A retrospective analysis of 140 patients responding to treatment with gefitinib revealed that the presence of BAC features ($p=0.005$) and being a never smoker ($p=0.007$) were the only independent predictors of response to gefitinib. These data suggest that EGFR inhibitor therapy is more active in BAC and in non-smokers.

However, currently, there are no selection criteria for determining which NSCLC patients will benefit from treatment with EGFR inhibitors such as gefitinib. Moreover, EGFR expression does not predict gefitinib sensitivity. Therefore, despite the correlation of tumor histology and smoking history with gefitinib response, it is of great importance to identify molecular molecules that influence gefitinib responsiveness, and to develop adjuvant treatments that enhance the response. To accomplish this goal, there is a need in the art to define critical aspects of EGFR signaling and to identify which molecules interact with the EGFR pathway to dictate responsiveness to EGFR inhibitors.

Description of the Invention

The present invention generally relates to the identification, provision and use of a panel of biomarkers that predict sensitivity or resistance to gefitinib, and products and processes related thereto. Specifically, the present inventors have used NSCLC cell lines with varying sensitivity to the EGFR inhibitor, gefitinib, to define the novel panel of biomarkers as described herein. In order to identify a marker panel that could be used for selection of NSCLC patients who will respond to gefitinib treatment, the inventors undertook preclinical *in vitro* studies using NSCLC cell lines. Based on the therapeutic response to gefitinib by using the IC 50 definition (i.e., the concentration of agent needed to kill 50% of the tumor cells in a cell culture), the present inventors have classified the cell lines as sensitive (IC5o<1 μM), resistant (>10 μM), or having intermediate sensitivity (1 μM <IC50< 10 μM) to gefitinib. The cell lines were characterized by gene microarray analysis (Affymetrix™ microarray HGU 133A, 22.000 genes). By comparing the gene microarray results from sensitive and resistant cell lines, the inventors have identified a panel of genes that can discriminate between sensitive and resistant cell lines. These biomarkers (i.e., the genes identified) will be of great clinical significance in selecting NSCLC patients/human tumors which will respond to this agent. The biomarkers identified by the present invention, and their expression levels in gefitinib sensitive and resistant cells, are listed in Table 1. It is to be understood that the present invention expressly covers additional genes that can be elucidated using substantially the same techniques used to identify the genes in Table 1 and that any of such additional genes can be used in the methods and products described herein for the genes in Table 1. Any reference to database

Accession numbers or other information regarding the genes in Table 1 is hereby incorporated by reference in its entirety.

In addition, the present invention will also be useful for the validation in other studies of the clinical significance of many of the specific biomarkers described herein, as well as the identification of preferred biomarker profiles, highly sensitive biomarkers, and targets for the design of novel therapeutic products and strategies. The biomarkers described herein are particularly useful in clinical practice to select the patients who will benefit most from EGFR inhibitor treatment and specifically, from gefitinib treatment.

The present inventors have already used the biomarkers described herein to identify specific targets for the further development of diagnostic and therapeutic approaches used in cancer. For example, E-cadherin is a calcium-dependent epithelial cell adhesion molecule that plays an important role in tumor invasiveness and metastatic potential. Reduced E-cadherin expression is associated with tumor cell dedifferentiation, advanced stage and reduced survival in patients with NSCLC. Using Western blot analysis, E-cadherin was expressed in three cell lines highly sensitive to gefitinib and its expression was lacking in six gefitinib resistant cell lines tested. High-density oligonucleotide microarrays were used to evaluate the gene expression pattern in 11 NSCLC cell lines and compared to gene expression in normal bronchial epithelium. E-cadherin expression was elevated in cell lines sensitive to gefitinib and downregulated in the resistant cell lines as compared to the normal bronchial epithelium. These results were confirmed with real-time RT-PCR. The expression of E-cadherin is regulated by zinc finger inhibitory proteins by the recruitment of histone deacetylases (HDAC). Using real-time RT-PCR, the expression of the two zinc-finger transcription factors, δEF1/ZEB1 and SIP1/ZEB2, involved in E-cadherin repression was evaluated. Results showed that both ZEB1 and SIP1 were expressed in gefitinib resistant cell lines and their expression was lacking in gefitinib sensitive cell lines. This expression pattern of ZEB1 and SIP1 was confirmed in the microarray analysis. The present inventors have also found that δEF1/ZEB1 and SIP1/ZEB2 may regulate Her3, which is an EGFR heterodimer. These data indicate that the expression of the zinc finger containing proteins SIP1 and ZEB1 may predict resistance to EGFR tyrosine kinase inhibitors and future studies directed at modulating the regulation of E-cadherin expression are expected to enhance the activity of EGFR inhibitors in NSCLC.

Finally, the present invention also relates to protein profiles which can discriminate between sensitive and resistant NSCLC tumors.

Prior to the present invention, to the best of the present inventors' knowledge, no single marker, or marker panel, has been demonstrated to be useful for selection of lung cancer patients who will benefit from EGFR inhibitor, and particularly, gefitinib, treatment. Nor are there any such markers (related to EGFR inhibitors) identified for other types of cancer.

Accordingly, using the gene expression profiles disclosed in Table 1 for gefitinib sensitive and resistant cells, one can rapidly, effectively and efficiently screen patients/human tumors for a level of sensitivity or resistance to gefitinib and also to other EGFR inhibitors having biological activity substantially similar to gefitinib (i.e., drugs having similar activities, gefitinib agonists and other derivatives). The results will allow for the identification of tumors/patients that are likely to benefit from administration of the drug and therefore, the genes are used to enhance the ability of the clinician to develop prognosis and treatment protocols for the individual patient. In addition, genes identified in Table 1 can be further validated as targets and then used in assays to identify therapeutic reagents useful for regulating the expression or activity of the target in a manner that improves sensitivity of a cell to gefitinib or analogs thereof. The knowledge provided from the expression profile of genes described herein and the identification additional genes using similar methods can also be used to identify the molecular mechanisms of EGFR inhibition, such knowledge being useful for the further development of new therapies and even analogs of gefitinib or other EGFR inhibitors with improved efficacies in cancer treatment. Moreover, given the knowledge of these genes, one can produce novel combinations of polynucleotides and/or antibodies and/or peptides for use in the various assays, diagnostic and/or therapeutic approaches described herein.

Various definitions and aspects of the invention will be described below, but the invention is not limited to any specific embodiments that may be used for illustrative or exemplary purposes.

According to the present invention, in general, the biological activity or biological action of a protein refers to any function(s) exhibited or performed by the protein that is ascribed to the naturally occurring form of the protein as measured or observed *in vivo* (i.e., in the natural physiological environment of the protein) or *in vitro* (i.e., under laboratory conditions). Modifications of a protein, such as in a homologue or mimetic (discussed below), may result in

proteins having the same biological activity as the naturally occurring protein, or in proteins having decreased or increased biological activity as compared to the naturally occurring protein. Modifications which result in a decrease in protein expression or a decrease in the activity of the protein, can be referred to as inactivation (complete or partial), down-regulation, or decreased action of a protein. Similarly, modifications which result in an increase in protein expression or an increase in the activity of the protein, can be referred to as amplification, overproduction, activation, enhancement, up-regulation or increased action of a protein.

According to the present invention, a "downstream gene" or "endpoint gene" is any gene, the expression of which is regulated (up or down) within a gefitinib sensitive or resistant cell. Selected sets of one, two, and preferably several or many of the genes (up to the number equivalent to all of the genes) of this invention can be used as end-points for rapid screening of patient cells for sensitivity or resistance to EGFR inhibitors such as gefitinib and for the other methods as described herein, including the identification of novel targets for the development of new cancer therapeutics.

As used herein, the term "homologue" is used to refer to a protein or peptide which differs from a naturally occurring protein or peptide (i.e., the "prototype" or "wild-type" protein) by minor modifications to the naturally occurring protein or peptide, but which maintains the basic protein and side chain structure of the naturally occurring form. Such changes include, but are not limited to: changes in one or a few amino acid side chains; changes one or a few amino acids, including deletions (e.g., a truncated version of the protein or peptide) insertions and/or substitutions; changes in stereochemistry of one or a few atoms; and/or minor derivatizations, including but not limited to: methylation, glycosylation, phosphorylation, acetylation, myristoylation, prenylation, palmitation, amidation and/or addition of glycosylphosphatidyl inositol. A homologue can have either enhanced, decreased, or substantially similar properties as compared to the naturally occurring protein or peptide. A homologue can include an agonist of a protein or an antagonist of a protein.

Homologues can be the result of natural allelic variation or natural mutation. A naturally occurring allelic variant of a nucleic acid encoding a protein is a gene that occurs at essentially the same locus (or loci) in the genome as the gene which encodes such protein, but which, due to natural variations caused by, for example, mutation or recombination, has a similar but not identical sequence. Allelic variants typically encode proteins having similar activity to that of the protein

encoded by the gene to which they are being compared. One class of allelic variants can encode the same protein but have different nucleic acid sequences due to the degeneracy of the genetic code. Allelic variants can also comprise alterations in the 5' or 3' untranslated regions of the gene (e.g., in regulatory control regions). Allelic variants are well known to those skilled in the art.

An agonist, as used herein, is a compound that is characterized by the ability to agonize (e.g., stimulate, induce, increase, enhance, or mimic) the biological activity of a naturally occurring or reference protein or compound. More particularly, an agonist can include, but is not limited to, a compound, protein, peptide, or nucleic acid that mimics or enhances the activity of the natural or reference compound, and includes any homologue, mimetic, or any suitable product of drug/compound/peptide design or selection which is characterized by its ability to agonize (e.g., stimulate, induce, increase, enhance) the biological activity of a naturally occurring or reference compound.

An antagonist refers to any compound which inhibits (e.g., antagonizes, reduces, decreases, blocks, reverses, or alters) the effect of a naturally occurring or reference compound as described above. More particularly, an antagonist is capable of acting in a manner relative to the activity of the reference compound, such that the biological activity of the natural or reference compound, is decreased in a manner that is antagonistic (e.g., against, a reversal of, contrary to) to the natural action of the reference compound. Such antagonists can include, but are not limited to, any compound, protein, peptide, or nucleic acid (including ribozymes and antisense) or product of drug/compound/peptide design or selection that provides the antagonistic effect.

Agonists and antagonists that are products of drug design can be produced using various methods known in the art. Various methods of drug design, useful to design mimetics or other compounds useful in the present invention are disclosed in Maulik et al., 1997, *Molecular Biotechnology: Therapeutic Applications and Strategies*, Wiley-Liss, Inc., which is incorporated herein by reference in its entirety. An agonist or antagonist can be obtained, for example, from molecular diversity strategies (a combination of related strategies allowing the rapid construction of large, chemically diverse molecule libraries), libraries of natural or synthetic compounds, in particular from chemical or combinatorial libraries (i.e., libraries of compounds that differ in sequence or size but that have the similar building blocks) or by rational, directed or random drug

design. See for example, Maulik et al., *supra*.

In a molecular diversity strategy, large compound libraries are synthesized, for example, from peptides, oligonucleotides, natural or synthetic steroidal compounds, carbohydrates and/or natural or synthetic organic and non-steroidal molecules, using biological, enzymatic and/or chemical approaches. The critical parameters in developing a molecular diversity strategy include subunit diversity, molecular size, and library diversity. The general goal of screening such libraries is to utilize sequential application of combinatorial selection to obtain high-affinity ligands for a desired target, and then to optimize the lead molecules by either random or directed design strategies. Methods of molecular diversity are described in detail in Maulik, et al., ibid.

As used herein, the term "mimetic" is used to refer to any natural or synthetic compound, peptide, oligonucleotide, carbohydrate and/or natural or synthetic organic molecule that is able to mimic the biological action of a naturally occurring or known synthetic compound.

As used herein, the term "putative regulatory compound" or "putative regulatory ligand" refers to compounds having an unknown regulatory activity, at least with respect to the ability of such compounds to regulate the expression or biological activity of a gene or protein encoded thereby, or to regulate sensitivity or resistance to an EGFR inhibitor as encompassed by the present invention.

In accordance with the present invention, an isolated polynucleotide, or an isolated nucleic acid molecule, is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. Polynucleotides useful in the plurality of polynucleotides of the present invention (described below) are typically a portion of a gene of the present invention that is suitable for use as a hybridization probe or PCR primer for the identification of a full-length gene (or portion thereof) in a given sample (e.g., a cell sample). An isolated nucleic acid molecule can include a gene or a portion of a gene (e.g., the regulatory region or promoter), for example, to produce a reporter construct according to the present invention. An isolated nucleic acid molecule that includes a gene

is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes naturally found on the same chromosome. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecule can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Although the phrase "nucleic acid molecule" primarily refers to the physical nucleic acid molecule and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the nucleic acid molecule, the two phrases can be used interchangeably, especially with respect to a nucleic acid molecule, or a nucleic acid sequence, being capable of encoding a protein. Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. If the polynucleotide is an oligonucleotide probe, the probe preferably ranges from about 5 to about 50 or about 500 nucleotides, more preferably from about 10 to about 40 nucleotides, and most preferably from about 15 to about 40 nucleotides in length.

An isolated protein, according to the present invention, is a protein (including a peptide) that has been removed from its natural milieu (i.e., that has been subject to human manipulation) and can include purified proteins, partially purified proteins, recombinantly produced proteins, and synthetically produced proteins, for example. As such, "isolated" does not reflect the extent to which the protein has been purified. An isolated protein useful as an antagonist or agonist according to the present invention can be isolated from its natural source, produced recombinantly or produced synthetically. Smaller peptides useful as regulatory peptides are typically produced synthetically by methods well known to those of skill in the art.

According to the present invention, the phrase "selectively binds to" refers to the ability of an antibody, antigen binding fragment or binding partner (antigen binding peptide) to preferentially bind to specified proteins. More specifically, the phrase "selectively binds" refers to the specific binding of one protein to another (*e.g.*, an antibody, fragment thereof, or binding partner to an antigen), wherein the level of binding, as measured by any standard assay (*e.g.*, an immunoassay), is statistically significantly higher than the background control for the assay. For example, when

performing an immunoassay, controls typically include a reaction well/tube that contain antibody or antigen binding fragment alone (*i.e.*, in the absence of antigen), wherein an amount of reactivity (*e.g.*, non-specific binding to the well) by the antibody or antigen binding fragment thereof in the absence of the antigen is considered to be background. Binding can be measured using a variety of methods standard in the art including enzyme immunoassays (*e.g.*, ELISA), immunoblot assays, etc.).

In some embodiments of the present invention, a compound is contacted with one or more nucleic acids or proteins. Such methods can include cell-based assays, or non-cell-based assay. In one embodiment, a target gene is expressed by a cell (i.e., a cell-based assay). In one embodiment, the conditions under which a cell expressing a target is contacted with a putative regulatory compound, such as by mixing, are conditions in which the expression or biological activity of the target (gene or protein encoded thereby) is not stimulated (activated) if essentially no regulatory compound is present. For example, such conditions include normal culture conditions in the absence of a known activating compound or other equivalent stimulus. The putative regulatory compound is then contacted with the cell. In this embodiment, the step of detecting is designed to indicate whether the putative regulatory compound alters the expression and/or biological activity of the gene or protein target as compared to in the absence of the putative regulatory compound (i.e., the background level).

In accordance with the present invention, a cell-based assay as described herein is conducted under conditions which are effective to screen for regulatory compounds or to profile gene expression as described in the methods of the present invention. Effective conditions include, but are not limited to, appropriate media, temperature, pH and oxygen conditions that permit the growth of the cell that expresses the receptor. An appropriate, or effective, medium is typically a solid or liquid medium comprising growth factors and assimilable carbon, nitrogen and phosphate sources, as well as appropriate salts, minerals, metals and other nutrients, such as vitamins. Culturing is carried out at a temperature, pH and oxygen content appropriate for the cell. Such culturing conditions are within the expertise of one of ordinary skill in the art.

Cells that are useful in the cell-based assays of the present invention include any cell that expresses a gene that is to be investigated as a target, or in the diagnostic assays described herein,

any cell that is isolated from a patient, including normal or malignant (tumor) cells.

According to the present invention, the method includes the step of detecting the expression of at least one, and preferably more than one, and most preferably, several, of the genes that have now been shown to be regulated differently in gefitinib-sensitive versus gefitinib-resistant cells by the present inventors. As used herein, the term "expression", when used in connection with detecting the expression of a gene, can refer to detecting transcription of the gene and/or to detecting translation of the gene. To detect expression of a gene refers to the act of actively determining whether a gene is expressed or not. This can include determining whether the gene expression is upregulated as compared to a control, downregulated as compared to a control, or unchanged as compared to a control. Therefore, the step of detecting expression does not require that expression of the gene actually is upregulated or downregulated, but rather, can also include detecting that the expression of the gene has not changed (i.e., detecting no expression of the gene or no change in expression of the gene).

The present method includes the step of detecting the expression of at least one gene set forth in Table 1. In a preferred embodiment, the step of detecting includes detecting the expression of at least 2 genes, and preferably at least 3 genes, and more preferably at least 4 genes, and more preferably at least 5 genes, and more preferably at least 6 genes, and more preferably at least 7 genes, and more preferably at least 8 genes, and more preferably at least 9 genes, and more preferably at least 10 genes, and more preferably at least 11 genes, and more preferably at least 12 genes, and more preferably at least 13 genes, and more preferably at least 14 genes, and more preferably at least 15 genes, and so on, in increments of one, up to detecting expression of all of the genes disclosed herein in Table 1. Preferably, larger numbers of genes in Table 1 are detected, as this will increase the sensitivity of the detection method. Analysis of a number of genes greater than 1 can be accomplished simultaneously, sequentially, or cumulatively.

In one aspect, it may be desirable to select those genes for detection that are particularly highly regulated in gefitinib-sensitive cells versus gefitinib-resistant cells in that they display the largest increases or decreases in expression levels. The detection of such genes can be advantageous because the endpoint may be more clear and require less quantitation. The relative expression levels of the genes identified in the present invention are listed in Table 1, and the genes are ranked in the

Table. Therefore, one can easily select subsets of particularly highly regulated genes, or subsets of genes based on some other desired characteristic to provide a more robust, sensitive, or selective assay. In one embodiment, one of skill in the art might choose to detect genes that exhibited a fold increase above background of at least 2. In another embodiment, one of skill in the art might choose to detect genes that exhibited a fold increase or decrease above background of at least 3, and in another embodiment at least 4, and in another embodiment at least 5, and in another embodiment at least 6, and in another embodiment at least 7, and in another embodiment at least 8, and in another embodiment at least 9, and in another embodiment at least 10 or higher fold changes. It is noted that fold increases or decreases are not typically compared from one gene to another, but with reference to the background level for that particular gene.

In one aspect of the method of the present invention, the step of detecting can include the detection of expression of one or more of the genes of this invention. Expression of transcripts and/or proteins is measured by any of a variety of known methods in the art. For RNA expression, methods include but are not limited to: extraction of cellular mRNA and northern blotting using labeled probes that hybridize to transcripts encoding all or part of one or more of the genes of this invention; amplification of mRNA expressed from one or more of the genes of this invention using gene-specific primers and reverse transcriptase - polymerase chain reaction, followed by quantitative detection of the product by any of a variety of means; extraction of total RNA from the cells, which is then labeled and used to probe cDNAs or oligonucleotides encoding all or part of the genes of this invention, arrayed on any of a variety of surfaces.

Methods to measure protein expression levels of selected genes of this invention, include, but are not limited to: western blotting, immunocytochemistry, flow cytometry or other immunologic-based assays; assays based on a property of the protein including but not limited to DNA binding, ligand binding, or interaction with other protein partners.

Nucleic acid arrays are particularly useful for detecting the expression of the genes of the present invention. The production and application of high-density arrays in gene expression monitoring have been disclosed previously in, for example, WO 97/10365; WO 92/10588; U.S. Patent No. 6,040,138; U.S. 5,445,934; or WO95/35505, all of which are incorporated herein by reference in their entireties. Also for examples of arrays, see Hacia *et al.* (1996) *Nature Genetics*

14:441-447; Lockhart *et al.* (1996) Nature Biotechnol. 14:1675-1680; and De Risi *et al.* (1996) *Nature Genetics* 14:457-460. In general, in an array, an oligonucleotide, a cDNA, or genomic DNA, that is a portion of a known gene occupies a known location on a substrate. A nucleic acid target sample is hybridized with an array of such oligonucleotides and then the amount of target nucleic acids hybridized to each probe in the array is quantified. One preferred quantifying method is to use confocal microscope and fluorescent labels. The *Affymetrix GeneChip™ Array* system (Affymetrix, Santa Clara, Calif.) and the *Atlas™ Human cDNA Expression Array* system are particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar systems or other effectively equivalent detection methods can also be used. In a particularly preferred embodiment, one can use the knowledge of the genes described herein to design novel arrays of polynucleotides, cDNAs or genomic DNAs for screening methods described herein. Such novel pluralities of polynucleotides are contemplated to be a part of the present invention and are described in detail below.

Suitable nucleic acid samples for screening on an array contain transcripts of interest or nucleic acids derived from the transcripts of interest. As used herein, a nucleic acid derived from a transcript refers to a nucleic acid for whose synthesis the mRNA transcript or a subsequence thereof has ultimately served as a template. Thus, a cDNA reverse transcribed from a transcript, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, etc., are all derived from the transcript and detection of such derived products is indicative of the presence and/or abundance of the original transcript in a sample. Thus, suitable samples include, but are not limited to, transcripts of the gene or genes, cDNA reverse transcribed from the transcript, cRNA transcribed from the cDNA, DNA amplified from the genes, RNA transcribed from amplified DNA, and the like. Preferably, the nucleic acids for screening are obtained from a homogenate of cells or tissues or other biological samples. Preferably, such sample is a total RNA preparation of a biological sample. More preferably in some embodiments, such a nucleic acid sample is the total mRNA isolated from a biological sample. Biological samples may be of any biological tissue or fluid or cells from any organism. Frequently the sample will be a "clinical sample" which is a sample derived from a patient, such as a lung tumor sample from a patient. Typical clinical samples include, but are not limited to, sputum, blood, blood cells (e.g.,

white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. Biological samples may also include sections of tissues, such as frozen sections or formalin fixed sections taken for histological purposes.

In one embodiment, it is desirable to amplify the nucleic acid sample prior to hybridization. One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids to achieve quantitative amplification. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. The high-density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid. Other suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al., Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), transcription amplification (Kwoh, et al., Proc. Natl. Acad. Sci. USA, 86: 1173 (1989)), and self-sustained sequence replication (Guatelli, et al, Proc. Nat. Acad. Sci. USA, 87: 1874 (1990)).

Nucleic acid hybridization simply involves contacting a probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing. As used herein, hybridization conditions refer to standard hybridization conditions under which nucleic acid molecules are used to identify similar nucleic acid molecules. Such standard conditions are disclosed, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety (see specifically, pages 9.31-9.62). In addition, formulae to calculate the appropriate hybridization and wash conditions to achieve hybridization permitting varying degrees of mismatch of nucleotides are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284; Meinkoth et al., *ibid.*, is incorporated by reference herein in its entirety. Nucleic acids that do not form hybrid duplexes are washed away from the hybridized

nucleic acids and the hybridized nucleic acids can then be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

High stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 90% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 10% or less mismatch of nucleotides). One of skill in the art can use the formulae in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284 (incorporated herein by reference in its entirety) to calculate the appropriate hybridization and wash conditions to achieve these particular levels of nucleotide mismatch. Such conditions will vary, depending on whether DNA:RNA or DNA:DNA hybrids are being formed. Calculated melting temperatures for DNA:DNA hybrids are 10°C less than for DNA:RNA hybrids. In particular embodiments, stringent hybridization conditions for DNA:DNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 20°C and about 35°C, more preferably, between about 28°C and about 40°C, and even more preferably, between about 35°C and about 45°C. In particular embodiments, stringent hybridization conditions for DNA:RNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na^+) at a temperature of between about 30°C and about 45°C, more preferably, between about 38°C and about 50°C, and even more preferably, between about 45°C and about 55°C. These values are based on calculations of a melting temperature for molecules larger than about 100 nucleotides, 0% formamide and a G + C content of about 40%. Alternatively, T_m can be calculated empirically as set forth in Sambrook et al., *supra*, pages 9.31 to 9.62.

The hybridized nucleic acids are detected by detecting one or more labels attached to the sample nucleic acids. The labels may be incorporated by any of a number of means well known to

those of skill in the art. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads.TM.), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

The term "quantifying" or "quantitating" when used in the context of quantifying transcription levels of a gene can refer to absolute or to relative quantification. Absolute quantification may be accomplished by inclusion of known concentration(s) of one or more target nucleic acids and referencing the hybridization intensity of unknowns with the known target nucleic acids (e.g. through generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication, transcription level.

In one aspect of the present method, *in vitro* cell based assays may be designed to screen for compounds that affect the regulation of genes at either the transcriptional or translational level. One, two or more promoters of the genes of this invention can be used to screen unknown compounds for activity on a given target. Promoters of the selected genes can be linked to any of several reporters (including but not limited to chloramphenicol acetyl transferase, or luciferase) that measure transcriptional read-out. The promoters can be tested as pure DNA, or as DNA bound to chromatin proteins.

In one aspect of the present method, the step of detecting can include detecting the expression

of one or more genes of the invention in intact animals or tissues obtained from such animals. Mammalian (i.e. mouse, rat, monkey) or non-mammalian (ie. chicken) species can be the test animals. Sample tissues from a patient can also be screened. The tissues to be surveyed can be either normal or malignant tissues. The presence and quantity of endogenous mRNA or protein expression of one or more of the genes of this invention can be measured in those tissues. The gene markers can be measured in tissues that are fresh, frozen, fixed or otherwise preserved. They can be measured in cytoplasmic or nuclear organ-, tissue- or cell-extracts; or in cell membranes including but not limited to plasma, cytoplasmic, mitochondrial, golgi or nuclear membranes; in the nuclear matrix; or in cellular organelles and their extracts including but not limited to ribosomes, nuclei, nucleoli, mitochondria, or golgi. Assays for endogenous expression of mRNAs or proteins encoded by the genes of this invention can be performed as described above. Alternatively, intact transgenic animals can be generated for screening for research or validation purposes.

Preferably, a gene identified as being upregulated or downregulated in a test cell according to the invention (including a sample tumor cell to be screened) is regulated in the same direction and to at least about 10%, and more preferably at least 20%, and more preferably at least 25%, and more preferably at least 30%, and more preferably at least 35%, and more preferably at least 40%, and more preferably at least 45%, and more preferably at least 50%, and preferably at least 55%, and more preferably at least 60%, and more preferably at least 65%, and more preferably at least 70%, and more preferably at least 75%, and more preferably at least 80%, and more preferably at least 85%, and more preferably at least 90%, and more preferably at least 95%, of the level of expression of the gene that is seen in established or confirmed gefitinib-sensitive or gefitinib-resistant cells. Statistical significance should be at least $p<0.05$.

It will be appreciated by those of skill in the art that differences between the expression of genes in sensitive versus resistant cells may be small or large. Some small differences may be very reproducible and therefore nonetheless useful. For other purposes, large differences may be desirable for ease of detection of the activity. It will be therefore appreciated that the exact boundary between what is called a positive result and a negative result can shift, depending on the goal of the screening assay and the genes to be screened. For some assays it may be useful to set threshold levels of change. One of skill in the art can readily determine the criteria for screening of cells given

the information provided herein.

The presence and quantity of each gene marker can be measured in primary tumors, metastatic tumors, locally recurring tumors, ductal carcinomas in situ, or other tumors. The markers can be measured in solid tumors that are fresh, frozen, fixed or otherwise preserved. They can be measured in cytoplasmic or nuclear tumor extracts; or in tumor membranes including but not limited to plasma, mitochondrial, golgi or nuclear membranes; in the nuclear matrix; or in tumor cell organelles and their extracts including but not limited to ribosomes, nuclei, mitochondria, golgi.

A profile of individual gene markers, including a matrix of two or more markers, can be generated by one or more of the methods described above. According to the present invention, a profile of the genes in a tissue sample refers to a reporting of the expression level of a given gene from Table 1, and includes a classification of the gene with regard to how the gene is regulated in gefitinib-sensitive versus gefitinib-resistant cells. The data can be reported as raw data, and/or statistically analyzed by any of a variety of methods, and/or combined with any other prognostic marker(s).

Another embodiment of the present invention relates to a plurality of polynucleotides for the detection of the expression of genes as described herein. The plurality of polynucleotides consists of polynucleotide probes that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes listed in Table 1 or otherwise identified as being useful according to the present invention, and is therefore distinguished from previously known nucleic acid arrays and primer sets. The plurality of polynucleotides within the above-limitation includes at least one or more, but is not limited to one or more, polynucleotide probes that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes identified by the present inventors and listed in Table 1.

In one embodiment, it is contemplated that additional genes that are not regulated differently in gefitinib-sensitive versus gefitinib-resistant cells can be added to the plurality of polynucleotides. Such genes would not be random genes, or large groups of unselected human genes, as are commercially available now, but rather, would be specifically selected to complement the sets of genes identified by the present invention. For example, one of skill in the art may wish to add to the above-described plurality of genes one or more genes that are of relevance because they are expressed by a particular tissue of interest (e.g., lung tissue), are associated with a particular disease

or condition of interest (e.g., NSCLC), or are associated with a particular cell, tissue or body function (e.g., angiogenesis). The development of additional pluralities of polynucleotides (and antibodies, as disclosed below), which include both the above-described plurality and such additional selected polynucleotides, are explicitly contemplated by the present invention.

According to the present invention, a plurality of polynucleotides refers to at least 2, and more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in increments of one, up to any suitable number of polynucleotides, including at least 100, 500, 1000, 10^4, 10^5, or at least 10^6 or more polynucleotides.

In one embodiment, the polynucleotide probes are conjugated to detectable markers. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads.TM.), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Preferably, the polynucleotide probes are immobilized on a substrate.

In one embodiment, the polynucleotide probes are hybridizable array elements in a microarray or high density array. Nucleic acid arrays are well known in the art and are described for use in comparing expression levels of particular genes of interest, for example, in U.S. Patent No. 6,177,248, which is incorporated herein by reference in its entirety. Nucleic acid arrays are suitable for quantifying a small variations in expression levels of a gene in the presence of a large population of heterogeneous nucleic acids. Knowing the identity of the genes of the present invention, nucleic acid arrays can be fabricated either by *de novo* synthesis on a substrate or by spotting or transporting nucleic acid sequences onto specific locations of substrate. Nucleic acids are purified and/or isolated from biological materials, such as a bacterial plasmid containing a cloned segment of sequence of interest. It is noted that all of the genes identified by the present invention have been previously

sequenced, at least in part, such that oligonucleotides suitable for the identification of such nucleic acids can be produced. The database accession number for each of the genes identified by the present inventors is provided in Table 1. Suitable nucleic acids are also produced by amplification of template, such as by polymerase chain reaction or in vitro transcription.

Synthesized oligonucleotide arrays are particularly preferred for this aspect of the invention. Oligonucleotide arrays have numerous advantages, as opposed to other methods, such as efficiency of production, reduced intra- and inter array variability, increased information content and high signal-to-noise ratio.

One of skill in the art will appreciate that an enormous number of array designs are suitable for the practice of this invention. An array will typically include a number of probes that specifically hybridize to the sequences of interest. In addition, in a preferred embodiment, the array will include one or more control probes. The high-density array chip includes "test probes." Test probes could be oligonucleotides that range from about 5 to about 45 or 5 to about 500 nucleotides, more preferably from about 10 to about 40 nucleotides and most preferably from about 15 to about 40 nucleotides in length. In other particularly preferred embodiments the probes are 20 or 25 nucleotides in length. In another preferred embodiments, test probes are double or single strand DNA sequences. DNA sequences are isolated or cloned from natural sources or amplified from natural sources using natural nucleic acids as templates, or produced synthetically. These probes have sequences complementary to particular subsequences of the genes whose expression they are designed to detect. Thus, the test probes are capable of specifically hybridizing to the target nucleic acid they are to detect.

Another embodiment of the present invention relates to a plurality of antibodies, or antigen binding fragments thereof, for the detection of the expression of genes according to the present invention. The plurality of antibodies, or antigen binding fragments thereof, consists of antibodies, or antigen binding fragments thereof, that selectively bind to proteins encoded by genes described herein. According to the present invention, a plurality of antibodies, or antigen binding fragments thereof, refers to at least 2, and more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in

increments of one, up to any suitable number of antibodies, or antigen binding fragments thereof, including at least 100, 500, or at least 1000 antibodies, or antigen binding fragments thereof.

The invention also extends to non-antibody polypeptides, sometimes referred to as binding partners or antigen binding peptides, that have been designed to bind specifically to, and either activate or inhibit as appropriate, a target protein. Examples of the design of such polypeptides, which possess a prescribed ligand specificity are given in Beste et al. (*Proc. Natl. Acad. Sci.* 96:1898-1903, 1999), incorporated herein by reference in its entirety.

Limited digestion of an immunoglobulin with a protease may produce two fragments. An antigen binding fragment is referred to as an Fab, an Fab', or an $F(ab')_2$ fragment. A fragment lacking the ability to bind to antigen is referred to as an Fc fragment. An Fab fragment comprises one arm of an immunoglobulin molecule containing a L chain ($V_L + C_L$ domains) paired with the V_H region and a portion of the C_H region (CH1 domain). An Fab' fragment corresponds to an Fab fragment with part of the hinge region attached to the CH1 domain. An $F(ab')_2$ fragment corresponds to two Fab' fragments that are normally covalently linked to each other through a di-sulfide bond, typically in the hinge regions.

Isolated antibodies of the present invention can include serum containing such antibodies, or antibodies that have been purified to varying degrees. Whole antibodies of the present invention can be polyclonal or monoclonal. Alternatively, functional equivalents of whole antibodies, such as antigen binding fragments in which one or more antibody domains are truncated or absent (e.g., Fv, Fab, Fab', or $F(ab)_2$ fragments), as well as genetically-engineered antibodies or antigen binding fragments thereof, including single chain antibodies or antibodies that can bind to more than one epitope (e.g., bi-specific antibodies), or antibodies that can bind to one or more different antigens (e.g., bi- or multi-specific antibodies), may also be employed in the invention.

Generally, in the production of an antibody, a suitable experimental animal, such as, for example, but not limited to, a rabbit, a sheep, a hamster, a guinea pig, a mouse, a rat, or a chicken, is exposed to an antigen against which an antibody is desired. Typically, an animal is immunized with an effective amount of antigen that is injected into the animal. An effective amount of antigen refers to an amount needed to induce antibody production by the animal. The animal's immune system is then allowed to respond over a pre-determined period of time. The immunization process

can be repeated until the immune system is found to be producing antibodies to the antigen. In order to obtain polyclonal antibodies specific for the antigen, serum is collected from the animal that contains the desired antibodies (or in the case of a chicken, antibody can be collected from the eggs). Such serum is useful as a reagent. Polyclonal antibodies can be further purified from the serum (or eggs) by, for example, treating the serum with ammonium sulfate.

Monoclonal antibodies may be produced according to the methodology of Kohler and Milstein (*Nature* 256:495-497, 1975). For example, B lymphocytes are recovered from the spleen (or any suitable tissue) of an immunized animal and then fused with myeloma cells to obtain a population of hybridoma cells capable of continual growth in suitable culture medium. Hybridomas producing the desired antibody are selected by testing the ability of the antibody produced by the hybridoma to bind to the desired antigen.

Finally, any of the genes of this invention, or their RNA or protein products, can serve as targets for therapeutic strategies. For example, neutralizing antibodies could be directed against one of the protein products of a selected gene, expressed on the surface of a tumor cell. Alternatively regulatory compounds that regulate (e.g., upregulate or downregulate) the expression and/or biological activity of a target gene (whether the product is intracellular, membrane or secreted), can be identified and/or designed using the genes described herein.

If a suitable therapeutic compound is identified using the methods and genes of the present invention, a composition can be formulated. A composition, and particularly a therapeutic composition, of the present invention generally includes the therapeutic compound and a carrier, and preferably, a pharmaceutically acceptable carrier. According to the present invention, a "pharmaceutically acceptable carrier" includes pharmaceutically acceptable excipients and/or pharmaceutically acceptable delivery vehicles, which are suitable for use in administration of the composition to a suitable *in vitro*, *ex vivo* or *in vivo* site. A suitable *in vitro*, *in vivo* or *ex vivo* site is preferably a tumor cell. In some embodiments, a suitable site for delivery is a site of inflammation, near the site of a tumor, or a site of any other disease or condition in which regulation of the genes identified herein can be beneficial. Preferred pharmaceutically acceptable carriers are capable of maintaining a compound, a protein, a peptide, nucleic acid molecule or mimetic (drug) according to the present invention in a form that, upon arrival of the compound, protein, peptide,

nucleic acid molecule or mimetic at the cell target in a culture or in patient, the compound, protein, peptide, nucleic acid molecule or mimetic is capable of interacting with its target.

Suitable excipients of the present invention include excipients or formularies that transport or help transport, but do not specifically target a composition to a cell (also referred to herein as non-targeting carriers). Examples of pharmaceutically acceptable excipients include, but are not limited to water, phosphate buffered saline, Ringer's solution, dextrose solution, serum-containing solutions, Hank's solution, other aqueous physiologically balanced solutions, oils, esters and glycols. Aqueous carriers can contain suitable auxiliary substances required to approximate the physiological conditions of the recipient, for example, by enhancing chemical stability and isotonicity.

Suitable auxiliary substances include, for example, sodium acetate, sodium chloride, sodium lactate, potassium chloride, calcium chloride, and other substances used to produce phosphate buffer, Tris buffer, and bicarbonate buffer. Auxiliary substances can also include preservatives, such as thimerosal, m- or o-cresol, formalin and benzol alcohol. Compositions of the present invention can be sterilized by conventional methods and/or lyophilized.

One type of pharmaceutically acceptable carrier includes a controlled release formulation that is capable of slowly releasing a composition of the present invention into a patient or culture. As used herein, a controlled release formulation comprises a compound of the present invention (e.g., a protein (including homologues), a drug, an antibody, a nucleic acid molecule, or a mimetic) in a controlled release vehicle. Suitable controlled release vehicles include, but are not limited to, biocompatible polymers, other polymeric matrices, capsules, microcapsules, microparticles, bolus preparations, osmotic pumps, diffusion devices, liposomes, lipospheres, and transdermal delivery systems. Other carriers of the present invention include liquids that, upon administration to a patient, form a solid or a gel *in situ*. Preferred carriers are also biodegradable (i.e., bioerodible). When the compound is a recombinant nucleic acid molecule, suitable delivery vehicles include, but are not limited to liposomes, viral vectors or other delivery vehicles, including ribozymes. Natural lipid-containing delivery vehicles include cells and cellular membranes. Artificial lipid-containing delivery vehicles include liposomes and micelles. A delivery vehicle of the present invention can be modified to target to a particular site in a patient, thereby targeting and making use of a compound of the present invention at that site. Suitable modifications include manipulating the chemical

formula of the lipid portion of the delivery vehicle and/or introducing into the vehicle a targeting agent capable of specifically targeting a delivery vehicle to a preferred site, for example, a preferred cell type. Other suitable delivery vehicles include gold particles, poly-L-lysine/DNA-molecular conjugates, and artificial chromosomes.

A pharmaceutically acceptable carrier which is capable of targeting is herein referred to as a "delivery vehicle." Delivery vehicles of the present invention are capable of delivering a composition of the present invention to a target site in a patient. A "target site" refers to a site in a patient to which one desires to deliver a composition. For example, a target site can be any cell which is targeted by direct injection or delivery using liposomes, viral vectors or other delivery vehicles, including ribozymes and antibodies. Examples of delivery vehicles include, but are not limited to, artificial and natural lipid-containing delivery vehicles, viral vectors, and ribozymes. Natural lipid-containing delivery vehicles include cells and cellular membranes. Artificial lipid-containing delivery vehicles include liposomes and micelles. A delivery vehicle of the present invention can be modified to target to a particular site in a subject, thereby targeting and making use of a compound of the present invention at that site. Suitable modifications include manipulating the chemical formula of the lipid portion of the delivery vehicle and/or introducing into the vehicle a compound capable of specifically targeting a delivery vehicle to a preferred site, for example, a preferred cell type. Specifically, targeting refers to causing a delivery vehicle to bind to a particular cell by the interaction of the compound in the vehicle to a molecule on the surface of the cell. Suitable targeting compounds include ligands capable of selectively (i.e., specifically) binding another molecule at a particular site. Examples of such ligands include antibodies, antigens, receptors and receptor ligands. Manipulating the chemical formula of the lipid portion of the delivery vehicle can modulate the extracellular or intracellular targeting of the delivery vehicle. For example, a chemical can be added to the lipid formula of a liposome that alters the charge of the lipid bilayer of the liposome so that the liposome fuses with particular cells having particular charge characteristics.

Another preferred delivery vehicle comprises a viral vector. A viral vector includes an isolated nucleic acid molecule useful in the present invention, in which the nucleic acid molecules are packaged in a viral coat that allows entrance of DNA into a cell. A number of viral vectors can

be used, including, but not limited to, those based on alphaviruses, poxviruses, adenoviruses, herpesviruses, lentiviruses, adeno-associated viruses and retroviruses.

A composition can be delivered to a cell culture or patient by any suitable method. Selection of such a method will vary with the type of compound being administered or delivered (i.e., compound, protein, peptide, nucleic acid molecule, or mimetic), the mode of delivery (i.e., *in vitro*, *in vivo*, *ex vivo*) and the goal to be achieved by administration/delivery of the compound or composition. According to the present invention, an effective administration protocol (i.e., administering a composition in an effective manner) comprises suitable dose parameters and modes of administration that result in delivery of a composition to a desired site (i.e., to a desired cell) and/or in the desired regulatory event.

Administration routes include *in vivo*, *in vitro* and *ex vivo* routes. *In vivo* routes include, but are not limited to, oral, nasal, intratracheal injection, inhaled, transdermal, rectal, and parenteral routes. Preferred parenteral routes can include, but are not limited to, subcutaneous, intradermal, intravenous, intramuscular and intraperitoneal routes. Intravenous, intraperitoneal, intradermal, subcutaneous and intramuscular administrations can be performed using methods standard in the art. Aerosol (inhalation) delivery can also be performed using methods standard in the art (see, for example, Stribling et al., *Proc. Natl. Acad. Sci. USA* 189:11277-11281, 1992, which is incorporated herein by reference in its entirety). Oral delivery can be performed by complexing a therapeutic composition of the present invention to a carrier capable of withstanding degradation by digestive enzymes in the gut of an animal. Examples of such carriers, include plastic capsules or tablets, such as those known in the art. Direct injection techniques are particularly useful for suppressing graft rejection by, for example, injecting the composition into the transplanted tissue, or for site-specific administration of a compound, such as at the site of a tumor. *Ex vivo* refers to performing part of the regulatory step outside of the patient, such as by transfecting a population of cells removed from a patient with a recombinant molecule comprising a nucleic acid sequence encoding a protein according to the present invention under conditions such that the recombinant molecule is subsequently expressed by the transfected cell, and returning the transfected cells to the patient. *In vitro* and *ex vivo* routes of administration of a composition to a culture of host cells can be accomplished by a method including, but not limited to, transfection, transformation, electroporation,

microinjection, lipofection, adsorption, protoplast fusion, use of protein carrying agents, use of ion carrying agents, use of detergents for cell permeabilization, and simply mixing (e.g., combining) a compound in culture with a target cell.

In the method of the present invention, a therapeutic compound, as well as compositions comprising such compounds, can be administered to any organism, and particularly, to any member of the Vertebrate class, Mammalia, including, without limitation, primates, rodents, livestock and domestic pets. Livestock include mammals to be consumed or that produce useful products (e.g., sheep for wool production). Preferred mammals to protect include humans. Typically, it is desirable to obtain a therapeutic benefit in a patient. A therapeutic benefit is not necessarily a cure for a particular disease or condition, but rather, preferably encompasses a result which can include alleviation of the disease or condition, elimination of the disease or condition, reduction of a symptom associated with the disease or condition, prevention or alleviation of a secondary disease or condition resulting from the occurrence of a primary disease or condition, and/or prevention of the disease or condition. As used herein, the phrase "protected from a disease" refers to reducing the symptoms of the disease; reducing the occurrence of the disease, and/or reducing the severity of the disease. Protecting a patient can refer to the ability of a composition of the present invention, when administered to a patient, to prevent a disease from occurring and/or to cure or to alleviate disease symptoms, signs or causes. As such, to protect a patient from a disease includes both preventing disease occurrence (prophylactic treatment) and treating a patient that has a disease (therapeutic treatment) to reduce the symptoms of the disease. A beneficial effect can easily be assessed by one of ordinary skill in the art and/or by a trained clinician who is treating the patient. The term, "disease" refers to any deviation from the normal health of a mammal and includes a state when disease symptoms are present, as well as conditions in which a deviation (e.g., infection, gene mutation, genetic defect, etc.) has occurred, but symptoms are not yet manifested.

Various aspects of the invention are described in the following examples; however, the following examples are provided for the purpose of illustration and are not intended to limit the scope of the present invention.

Example 1

Methods: Gefitinib sensitivity was determined in 18 NSCLC cell lines using MTT assays. Cell lines were classified as gefitinib sensitive (IC< 1μM), resistant (IC50>10μM) or intermediate sensitivity (IC50>1,<10μM). Oligonucleotide gene arrays (Affymetrix® HGU 133A, 22.000 genes) were done on 10 cell lines. Three distinct filtration and normalization algorithms to process the expression data were used, and a list of genes were generated that were both statistically significant (unadjusted p=0.001 cutoff) and corrected for false positive occurrence. This approach was used in combination with 5 distinct machine learning algorithms used to build a test set for predictor genes that were successful for 100% of the test cases. The best discriminators (>3 fold difference in expression between sensitive and resistant cell lines) were selected for Real-time RT-PCR.

Results: A list of 144/210 genes was generated initially from the Affymetric array analysis. By using the mathematical algorithm 14 different candidate genes were selected for RT-PCR. Twelve of the 14 genes were verified to discriminate between sensitive and resistant cell lines by Real-time RT-PCR.

Conclusion: Based on NSCLC cell line studies it was possible to identify 12 different genes, which strongly discriminated gefitinib (Iressa) sensitive cell lines from the resistant ones. The genes are ranked in Table 1, with the top 12 genes listed first. This entire biomarker panel is of significant value for selecting NSCLC patients for gefitinib treatment.

Example 2

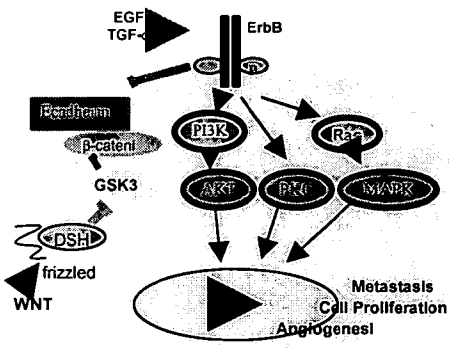
The attached report, labeled as "Example 2" describes the identification and further investigation of a target gene identified using the gene expression profile disclosed herein.

While various embodiments of the present invention have been described in detail, it is apparent that modifications and adaptations of those embodiments will occur to those skilled in the art. It is to be expressly understood, however, that such modifications and adaptations are within the scope of the present invention, as set forth in the following claims.

Example 2

In this study the present inventors describe research to examine the influence of E-cadherin-regulatory molecules on non-small cell lung cancer (NSCLC) response to EGF receptor (EGFR) inhibitors. The EGFR, a member of the erbB family of tyrosine kinases (erbB1-4) plays a major role in transmitting stimuli that lead to NSCLC cellular proliferation and survival. EGFR, highly expressed in NSCLC, is a primary target for NSCLC therapeutic intervention. In clinical trials, 11-20% of patients with chemo-refractory advanced stage NSCLC responded to treatment with EGFR inhibitors such as gefitinib (Iressa®, ZD1839). Currently, there are no markers that predict which patients will respond to treatment. NSCLC patients with poor survival have decreased expression of E-cadherin, a cell adhesion molecule. E-cadherin expression is regulated by the wnt pathway and by zinc finger transcription factors including δEF1/ZEB1 and SIP1/ZEB2. Higher levels of protein expression of E-cadherin were detected in gefitinib sensitive NSCLC cell lines and expression was absent in gefitinib resistant lines. Conversely, expression of the E-cadherin inhibitors ZEB1 and SIP1 was higher in gefitinib resistant cell lines. The Hypothesis of this project is that expression of E-cadherin and its regulatory molecules predict response to EGFR inhibitors, and modulating E-cadherin regulatory proteins may augment response to EGFR inhibitors in non-small cell lung cancer.

Chemotherapy is the mainstay of treatment for lung cancer, the leading cause of cancer deaths in men and women in the US and though out the world[1]. However less than a third of patients with advanced stages of NSCLC respond to the best two-chemotherapy drug combination.[2] Therefore novel agents that target cancer specific biological pathways are needed. The EGFR is one of the most appealing targets for novel therapies. EGFR plays a major role in transmitting stimuli that lead to proliferation, growth and survival of various cancer types including NSCLC. Ligand binding to the EGFR receptor leads to homo- or heterodimerization of EGFR with other ErbB receptors.[3] EGFR is overexpressed in a large proportion of invasive NSCLC and in premalignant bronchial lesions. Activation of the EGFR receptor leads to simultaneous activation of several signaling cascades including the MAPK pathway, the protein kinase C (PKC) pathway and the PI(3)K-activated AKT pathway (Figure 1). EGFR signaling translated in the nucleus leads to cancer cell proliferation and survival.[3]



In phase II clinical trials, 11-20% of patients with chemo-refractory advanced stage NSCLC responded to treatment with the EGFR tyrosine kinase inhibitor gefitinib[4,5]. A trial evaluating the activity of the EGFR inhibitor, erlotinib (Tarceva®, OSI-774) has been completed and the results will be reported in the near future. A retrospective analysis of 140 patients responding to treatment with gefitinib revealed that the presence of BAC features (p=0.005) and being a never smoker (p=0.007) were the only independent predictors of response to gefitinib[6]. These data suggest that EGFR inhibitor therapy is more active in BAC and in non-smokers.

Despite the correlation of tumor histology and smoking history with gefitinib response, it is of great importance to identify molecular molecules that influence gefitinib responsiveness, and to develop adjuvant treatments that enhance the response. To accomplish this goal, it is

critical that we understand the aspects of EGFR signaling and which molecules interacting with the EGFR pathway dictate responsiveness.

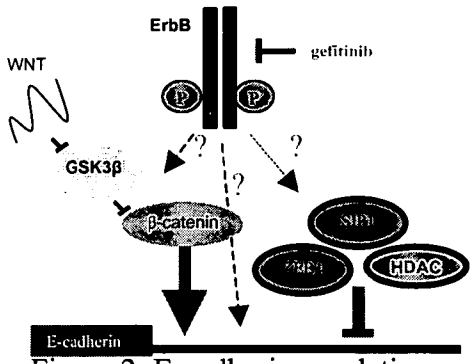


Figure 2: E-cadherin regulation

E-cadherin, a calcium-dependent epithelial cell adhesion molecule, plays an important role in tumor invasiveness and metastatic potential[7-10]. Reduced E-cadherin expression is associated with tumor cell dedifferentiation, advanced stage and reduced survival in patients with NSCLC[11,12]. At the transcriptional level, the wnt/β-catenin signaling pathway regulates DE-cadherin expression[13]. Our group recently reported that, inhibition of GSK3β, involved in the proteasomal degradation of β-catenin, lead E-cadherin upregulation[14] (Figure 2). E-cadherin transcription is also regulated by zinc finger transcription factors including, Snail, Slug, ZEB1 and SIP1[15-17]. They repress E-cadherin expression by binding to its promoter and recruiting HDAC[17](Figure 2). We recently reported that inhibiting the ZEB1 or HDAC expression lead to upregulation of E-cadherin in NSCLC cell lines[14].

We used NSCLC cell lines to: (1) evaluate the growth inhibitory properties of EGFR inhibitors by MTT assays, (2) to identify molecular molecules through DNA microarrays and westerns that predict response to EGFR inhibitors and (3) to design combination therapies that enhance the effect of the EGFR inhibitors. Cell lines were screened for expression of members of the EGFR and Wnt signaling pathways. We found that E-cadherin expression is lacking in gefitinib resistant cell lines and activated in gefitinib sensitive lines. Therefore we investigated the expression of zinc finger transcription factors involved in E-cadherin repression. We found that gefitinib resistant lines have high ZEB1 and/or SIP1 expression, and expression is lacking in gefitinib-sensitive lines.

Our hypothesis is that SIP1 and ZEB1 expression predicts EGFR tyrosine kinase inhibitors resistance and that modulating the molecular mechanism that regulate E-cadherin expression will enhance sensitivity to EGFR inhibitors. I will test this hypothesis by manipulating E-cadherin expression and measuring the effect on response to gefitinib. Results of this work will be evaluated in clinical trials in patients with NSCLC.

Results

I. EGFR, pEGFR, Her2, ErbB3 and Erb4 expression in NSCLC: EGFR, Her-2 and ErbB3 cell surface expression was evaluated using flow cytometry (Table 1). The majority of NSCLC cell lines (15/18) had a high percentage of EGFR positive cells and three had low or now EGFR expression. The two BAC cell lines, H322 and H358, had high expression of EGFR and Her2.

The presence of phosphorylated EGFR (pEGFR) vs EGFR was evaluated by Western blotting in 18 NSCLC cell lines (Figure 3, shows 15 cell lines). EGFR was detected in the majority of NSCLC cell lines whereas only a subset of these cell lines had (pEGFR).

Cell Line	FACS %EGFR/ MFI	FACS %Her2 MFI	FACS %ErbB3/ MFI	IC50 uM ZD1839
Adenocarcinoma				
Calu3	98%/8.9	100/37	32/4.3	<1
Colo699	0/0	0/0	57/2.3	4.1
H125	100/34	91/2.8	0/0	4.7
H2122	94/5.1	73/4	80/5	4.8
H1435	98/14	ND	94/6.4	7.6
A549	99/14	72/2.4	54/3.5	8.4
H441	78/6.9	79/2.6	0/0	11.7
H1648	98/5.7	78/2.7	0/0	11.5
Bronchoalveolar				
H322	100/16	98.5.5	ND	<1
H358	ND	ND	ND	<1
Squamous Cell				
NE18	100/16	98/3.3	35/5.7	8
H1703	99/15	65/2.6	0/0	9.3
H157	93/13	62/1.8	0/0	10.1
H520	0/0	0/0	0/0	10.3
H1264	100/14	43/1.9	0/0	10.2
Large Cell				
H1334	100/23	74/3.2	99/10	3.8
H460	37/1.9	57/1.4	0/0	9.9

Table 1: EGFR and Her2 expression by flow cytometry and immuno-fluorescence (MFI), and growth inhibition by the EGFR inhibitor

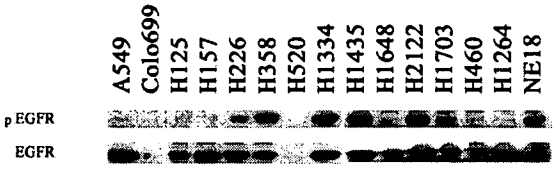


Figure 3. Expression of EGFR and phosphorylated EGFR in NSCLC cell lines.

II. Effects of EGFR inhibitors on human lung cancer cells growth: The growth inhibitory effect of gefitinib, on 18 NSCLC cell lines was evaluated using the MTT assay (Table 1). There was no correlation between the EGFR expression and gefitinib response. We evaluated the change pEGFR following gefitinib treatment in two sensitive cell lines, H1334 and H322, and two resistant cell lines, H1264 and H1648 (Figure 4). Gefitinib inhibited the phosphorylated "active" form of EGFR in sensitive cell lines.

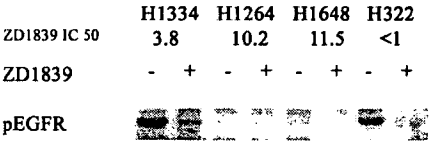


Figure 4: ZD1839 downregulates pEGF in sensitive NSCLC cell lines.

Based on the *in vitro* experiments, athymic nude mice bearing human NSCLC xenografts were treated with EGFR inhibitors ZD1839 or C225. Growth delay was evident in tumors after treatment with either agent (Figure 5).

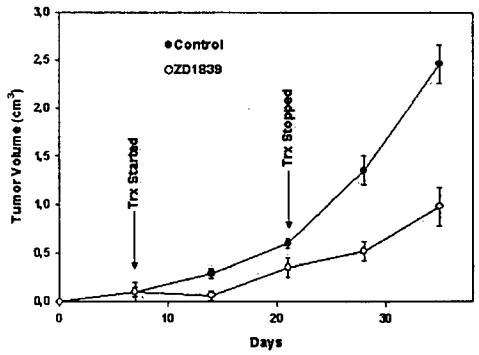


Figure 5: Effects of gefitinib on A549 NSCLC xenografts

III. E-cadherin, SIP1 and ZEB1 in NSCLC cell lines using microarray and RT-PCR and western blotting. High density oligonucleotide microarray (IOAM) analysis of gene expression levels of selected genes was developed from 11 NSCLC cell lines. These cell lines included 2 gefitinib sensitive lines (IC50<1μM), 5 gefitinib resistant lines (IC50≥ 10μM), and 4 lines with intermediate sensitivity (IC50 >1μM <10μM). The expression of E-cadherin, SIP1 and ZEB 1 was evaluated and compared to their expression in normal bronchial epithelium using the Gene Spring program (Figure 6).

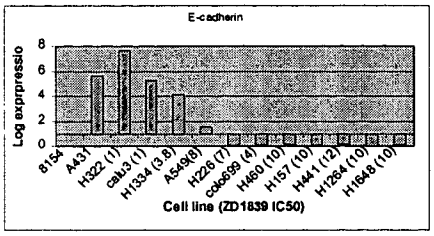


Figure 6: Expression of E-cadherin, i NSCLC cell lines using GeneSprin analysis of microarrays. Expression i compared in cell lines to their expressio

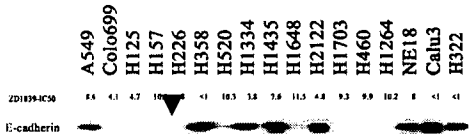


Figure 7: Western blott analysis of E-cadherin expression in NSCLC cell lines

E-cadherin expression was more pronounced in gefitinib sensitive lines absent in gefitinib resistant lines. This expression pattern was confirmed using western blotting and real time PCR (RT-PCR) (Figure 7).

As discussed above, regulation of E-cadherin expression involves the zinc finger transcription factors ZEB1 and SIP1. Expression of both transcription factors was evaluated using real time RT-PCR. ZEB 1 and SIP 1 were expressed in the gefitinib resistant lines and absent in the gefitinib sensitive lines (Figure 8). We also evaluated the expression of Slug, Snail, Wnt7a, β-catenin, γ-catein, α-catenin and GSK3β using Western blot analysis or RT-PCR. None of theses molecules had a differential pattern of expression in the NSCLC lines (data not shown).

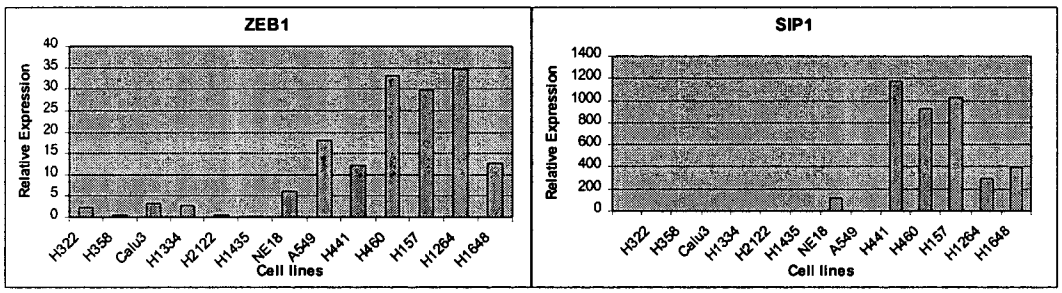


Figure 8: Real time RT-PCR analysis of ZEB1 and SIP1 expression in NSCLC cell lines.

In summary: There was no correlation between gefitinib sensitivity and EGFR expression. E-cadherin was detected preferentially in gefitinib sensitive lines. Conversely, the zinc finger transcription factors, ZEB1 and SIP1, involved in E-cadherin inhibition were expressed in gefitinib resistant lines and absent in gefitinib sensitive lines.

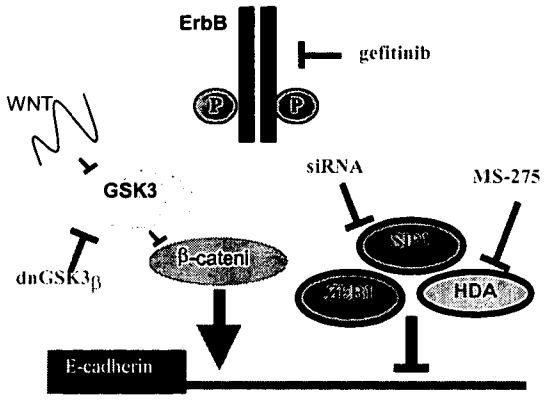
Future Research Design and Methods

Aim 1. Evaluate the detrimental effect of the zinc finger proteins ZEB1 and SIP1 on NSCLC cell lines sensitivity to EGFR inhibitors.

A. adenoviral constracts containing ZEB1 or SIP1 will be used to overexpress these proteins in gefitinib sensitive cell lines. MTT assay will assess changes in gefitinib sensitivity.
B. Stably transfected ZEB1 and SIP1 cell lines and untrasfected cell lines will be implanted into nude mice. Transplanted mice will be treated with gefitinib and response will be compared between the two groups.

Aim 2. Determine the molecular mechanisms that improve the response to EGFR inhibitors in NSCLC cell lines *in vitro* and *in vivo*.

2A. *Examining the effect of "silencing" the E-cadherin transcriptional repressors, SIP1 and ZEB1, on NSCLC cell lines response to ZD1839:*



To directly examine the role of the zinc-finger transcription factors, SIP1 and ZEB1, on gefitinib responsive lines, we will develop and test the effect of siRNA. siRNA will be prepared for different regions of SIP1 and ZEB1 using the silencer kit from Dharmacon (Colorado). Their efficacy will be tested by RT-PCR. The most effective siRNA for SIP1 and ZEB1 will be introduced, individually or in combination, into gefitinib resistant lines. The effect of these siRNAs on gefitinib responsiveness will be evaluated by MTT assay. ZEB1 antibody (Santa Cruz, California) and SIP1 antibody (a gift from Dr. van Grunsven) will be used to evaluate the efficacy of RNA inhibition.

2B. *Examining the effect of inhibiting GSK3β on gefitinib response in NSCLC cell lines:*
GSK3β phosphorylate β-catenin leading to its ubiqitination and destruction. GSK3β inhibitors, such as lithium, increased E-cadherin expression in NSCLC cell lines[14]. We will inhibit GSK3β function with an adenovirus (pAdTrack-CMV) encoding a dominant-negative GSK3β (dnGSK3β)(a gift from Dr. Woodgett). To determine the effectiveness of this dnGSK3 we will evaluate the expression of non-phosphorylated β-catenin and E-cadherin by western blot. NSCLC cell lines stably transfected with the dnGSK3β construct will be generated. The effect of inhibiting GSK3β on NSCLC cell lines response to gefitinib will be evaluated using MTT assays.

2C. *Evaluating the effect of E-cadherin on gefitinib sensitivity:*
Resistant NSCLC lines will be transfected with E-cadherin encoding constructs. Changes in NSCLC cell lines response to gefitinib will be assessed by MTT assay. Gefitinib-sensitive lines that express E-cadherin will be treated with an E-cadherin antibody (Zymed) and the effect on gefitinib responsiveness assessed by MTT assay. This will answer the question whether expression of E-cadherin itself is sufficient to determine gefitinib sensitivity or if sensitivity is a reflexion of events occurring upstream of it.

2D. *Augmenting the effect of gefitinib responsiveness on NSCLC cell lines in vivo:*
Based on finding from the above *in vitro* experiments, the best treatment that enhances gefitinib sensitivity in NSCLC cell lines will be selected for in vivo experiments in nude mice. Previously we showed an inhibitory effect of gefitinib alone on NSCLC xenografts growth (see above). We will test the combination of gefitinib with one of the above-evaluated interventions in athymic nude mice bearing human NSCLC xenografts in collaboration with Dr. Daniel Chan.
E-cadherin inducible cell lines from the in vitro experiments will be inoculated subcutaneously in nude mice. Mice will be treated with gefitinib with and without the agent that improved the gefitinib sensitivity. The two groups will be evaluated for differences in tumor growth inhibition. Expression of E-cadherin, SIP1 and ZEB1 will be evaluated both prior to and post-treatment by real-time RT-PCR and immunohistochemistry. ZEB1 antibody (Santa Cruz, California) and SIP1 antibody (a gift from Dr. van Grunsven) will be used in the immunohistochemistry. However, new antibodies will be generated if the above antibodies were not effective at detecting proteins in the IHC assays.

In the experimental design outlined above we hope to dissect out the events leading to gefitinib resistance develop treatment modifications that bypass resistance.

Aim III. Conduct clinical trial evaluating the correlation between ZEB1, SIP1 and resitance to EGFR inhibitors in patients with NSCLC.

This trial will enroll 29 patients with diagnosis proven Stage I/II NSCLC. Patients will be treated with GW572016 (a dual EGFR and Her2 inhibitor, GalaxoSmithKline Pharmaceuticals) for 4 weeks. Surgical removal of the tumor will be done at the end of this treatment period. Using RT-PCR analysis and immunohistochemistry, samples will be evaluated before and after treatment for ZEB1 and SIP1 expression. Correlation will be made between ZEB1 and SIP1 expression and EGFR inhibitor effectiveness.

Exemplary Claims:

1. A method to select a cancer patient who is predicted to benefit from therapeutic administration of gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib, comprising:

a) providing a sample of tumor cells from a patient to be tested;

b) detecting in the sample the expression of at least one gene selected from the genes listed in Table 1;

c) comparing the level of expression of the gene or genes detected in the patient sample to the level of expression of the gene in Table 1 that has been correlated with sensitivity and/or resistance to gefitinib;

d) selecting the patient as being predicted to benefit from therapeutic administration of gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib, if the expression of the gene or genes in the patient's tumor cells is statistically more similar to the expression levels of the gene or genes as has been correlated with sensitivity to gefitinib.

2. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least two genes from Table 1.

3. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least three genes from Table 1.

4. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least four genes from Table 1.

5. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least five genes from Table 1.

6. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least 10 genes from Table 1.

7. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least 25 genes from Table 1.

8. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least 50 genes from in Table 1.

9. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least 100 genes from Table 1.

10. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of at least 150 genes in Table 1.

11. The method of Claim 1, wherein the step (b) of detecting comprises detecting expression of substantially all of the genes in Table 1.

12. The method of any one of Claims 1-11, wherein expression of the gene or genes is detected by measuring amounts of transcripts of the gene in the tumor cells.

13. The method of any one of Claims 1-11, wherein expression of the gene or genes is detected by detecting hybridization of at least a portion of the gene or a transcript thereof to a nucleic acid molecule comprising a portion of the gene or a transcript thereof in a nucleic acid array.

14. The method of any one of Claims 1-11, wherein expression of the gene is detected by detecting the production of a protein encoded by the gene.

15. The method of any one of Claims 1-14, comprising detecting expression of: at least one gene selected from the group consisting of: E-cadherin, ErbB3, ZEB1 and SIP1.

16. The method of Claim 1, comprising comparing the expression of the gene or genes to expression of the gene or genes in a cell from a non-cancerous cell of the same type.

17. The method of Claim 1, comprising comparing the expression of the gene or genes to expression of the gene or genes in an autologous, non-cancerous cell from the patient.

18. The method of Claim 1, comprising comparing the expression of the gene or genes to expression of the gene or genes in a control cell that is resistant to gefitinib.

19. The method of Claim 1, comprising comparing the expression of the gene or genes to expression of the gene or genes in a control cell that is sensitive to gefitinib.

20. The method of Claim 1, wherein control expression levels of the gene or genes that has been correlated with sensitivity and/or resistance to gefitinib has been predetermined.

21. A method to identify molecules that interact with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors, comprising:

a) providing a sample of cells that are sensitive or resistant to treatment with gefitinib;

b) detecting the expression of at least one gene in the gefitinib-sensitive cells as compared to the level of expression of the gene or genes in the gefitinib-resistant cells;

c) identifying a gene or genes having a level of expression in gefitinib-sensitive cells that is statistically significantly different than the level of expression of the gene or genes gefitinib-resistant cells as potentially being a molecule that interacts with the EGFR pathway to allow or enhance responsiveness to EGFR inhibitors.

22. A plurality of polynucleotides for the detection of the expression of genes that are indicative of sensitivity or resistance to gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib;

wherein the plurality of polynucleotides consists of polynucleotide probes that are complementary to RNA transcripts, or nucleotides derived therefrom, of one or more genes listed in Table 1.

23. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least two of the genes from Table 1.

24. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least 5 of the genes from Table 1.

25. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least 10 of the genes from Table 1.

26. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least 25 of the genes from Table 1.

27. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least 50 of the genes from Table 1.

28. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of at least 100 of the genes from Table 1.

29. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides

consists of at least 150 of the genes from Table 1.

30. The plurality of polynucleotides of Claim 22, wherein the plurality of polynucleotides consists of all of the genes from Table 1.

31. The plurality of polynucleotides of any one of Claims 22-31, wherein said polynucleotide probes are immobilized on a substrate.

32. The plurality of polynucleotides of any one of Claims 22-31, wherein said polynucleotide probes are hybridizable array elements in a microarray.

33. The plurality of polynucleotides of any one of Claims 22-31, wherein said polynucleotide probes are conjugated to detectable markers.

34. A plurality of polynucleotides for the detection of the expression of genes that are indicative of sensitivity or resistance to gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib;

wherein the plurality of polynucleotides consists of polynucleotide probes that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes that are regulated in gefitinib-sensitive tumor cells as compared to gefitinib-resistant cells.

35. A plurality of antibodies, antigen binding fragments thereof, or antigen binding peptides, for the detection of the expression of genes that are indicative of sensitivity or resistance to gefitinib, an agonist thereof, or a drug having substantially similar biological activity as gefitinib;

wherein said plurality of antibodies, antigen binding fragments thereof, or antigen binding peptides consists of antibodies, antigen binding fragments thereof, or antigen binding peptides, each of which selectively binds to a protein encoded by a gene in Table 1.

36. A method to identify a compound with the potential to enhance the efficacy of EGFR inhibitors, comprising:

a) contacting a test compound with a cell that expresses a gene selected from any one or more of the genes identified in Table 1;

b) identifying compounds that increase the expression or activity of genes in Table 1 or the proteins encoded thereby that are correlated with sensitivity to gefitinib, or that decrease the expression or activity of genes in Table 1 or the proteins encoded thereby that are correlated with resistance to gefitinib, as compounds with potential to enhance the

efficacy of EGFR inhibitors.

37. The method of Claim 36, wherein the cell expresses a gene encoding E-cadherin or ErbB3, and wherein step (b) comprises identifying compounds that increase the expression or activity of E-cadherin or ErbB3 or the gene encoding E-cadherin or ErbB3.

38. The method of Claim 36, wherein the cell expresses a gene encoding ZEB1 and SIP1, wherein step (b) comprises identifying compounds that decrease the expression or activity ZEB1 or SIP1 or the gene encoding ZEB1 or SIP1.

39. A method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound identified by the method of Claim 36.

40. A method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound that upregulates the expression or activity of E-cadherin or ErbB3 or the gene encoding E-cadherin or ErbB3 in the tumor cells of the patient.

41. A method to treat a patient with a cancer, comprising administering to the patient a therapeutic composition comprising a compound that downregulates the expression of ZEB1 or SIP1 or the gene encoding ZEB1 or SIP1 in the tumor cells of the patient.

TABLE 1

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
1	p < 0.000001	2.771	236.088	228067_at	AW249666	Hs.136732	na	similar to RIKEN cDNA 2010300C02 gene
2	3.00E-06	81.404	946.766	203585_at	NM_007150	Hs.16622	ZNF185	zinc finger protein 185 (LIM domain)
3	4.00E-06	24.781	3842.474	203180_at	NM_000693	Hs.75746	ALDH1A3	aldehyde dehydrogenase 1 family, member A3
4	5.00E-06	7.236	115.101	215923_s_at	AK023421	Hs.309029	TIC	SEC7 homolog
5	6.00E-06	69.984	485.886	221610_s_at	BC000795	Hs.194385	STAP2	signal-transducing adaptor protein-2
6	6.00E-06	42.894	352.896	205455_at	NM_002447	Hs.2942	MST1R	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
7	1.20E-05	254.925	848.712	224800_at	AK022888.1	Hs.44743	WDFY1	WD repeat and FYVE domain containing 1
8	1.40E-05	10.541	161.391	211712_s_at	BC005830	Hs.430324	ANXA9	annexin A9
9	1.50E-05	41.259	675.598	202454_s_at	NM_001982	Hs.306251	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)
10	1.60E-05	500.835	1695.862	201778_s_at	NM_014774	Hs.269902	KIAA0494	KIAA0494 gene product
11	1.80E-05	92.824	227.621	206833_s_at	NM_001108	Hs.433071	ACYP2	acylphosphatase 2, muscle type
12	2.00E-05	83.346	307.554	218501_at	NM_019555	Hs.25951	ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3
13	2.20E-05	4.273	84.767	229245_at	AA535361	Hs.343666	LOC149267	hypothetical protein BC010522
14	3.70E-05	65.841	2182.528	202295_s_at	NM_004390	Hs.114931	CTSH	cathepsin H
15	4.50E-05	17.528	600.833	205780_at	NM_001197	Hs.155419	BIK	BCL2-interacting killer (apoptosis inducing)

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
16	4.80E-05	230.096	691.534	200864_s_at	NM_004663	Hs.75618	RAB11A	RAB11A, member RAS oncogene family
17	4.80E-05	7.3	816.641	201288_at	NM_001175	Hs.292738	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta
18	5.20E-05	5.326	431.331	222546_s_at	AW204755	Hs.55016	EPS8L2	EPS8-like 2
19	5.40E-05	2.605	156.469	232165_at	AL137725.1	Hs.200412	EPPK1	epiplakin 1
20	5.60E-05	154.836	502.789	209193_at	M24779	Hs.81170	PIM1	pim-1 oncogene
21	5.90E-05	661.21	4114.672	200799_at	NM_005345	Hs.75452	HSPA1A	heat shock 70kDa protein 1A
22	5.90E-05	3.663	103.179	206482_at	NM_005975	Hs.51133	PTK6	PTK6 protein tyrosine kinase 6
23	6.90E-05	175.85	490.125	212631_at	AI566082	Hs.152335	na	LOC154084
24	6.90E-05	15.02	70.016	223796_at	AF333769.1	Hs.212839	CASPR3	cell recognition molecule CASPR3
25	7.10E-05	56.554	167.655	212700_x_at	AJ002220	Hs.420584	KIAA0356	KIAA0356 gene product
26	7.40E-05	8.503	117.169	225757_s_at	AU147564	Hs.301478	CLMN	calmin (calponin-like, transmembrane)
27	7.50E-05	7.712	228.328	223232_s_at	AI768894	Hs.18376	CGN	cingulin
28	7.50E-05	3.296	46.735	228440_at	BE550153	Hs.119316	PET112L	PET112-like (yeast)
29	7.50E-05	82.366	204.715	209588_at	AL530874	Hs.125124	EPHB2	EphB2
30	7.90E-05	313.279	826.145	40359_at	M91083	Hs.72925	C11orf13	chromosome 11 open reading frame 13
31	8.00E-05	421.223	2112.539	1007_s_at	U48705mRNA	Hs.423573	DDR1	discoidin domain receptor family, member 1
32	8.10E-05	37.124	157.883	207747_s_at	NM_018110	Hs.279832	DOK4	docking protein 4
33	8.40E-05	73.28	581.749	212096_s_at	AL096842	Hs.7946	MTSG1	mitochondrial tumor suppressor gene 1
34	8.60E-05	628.105	1346.483	217873_at	NM_016289	Hs.6406	MO25	MO25 protein
35	8.60E-05	1560.6	2742.098	217827_s_at	NM_016630	Hs.242458	ACP33	acid cluster protein 33
36	8.90E-05	313.108	542.212	225192_at	AA044726	Hs.408319	MGC33215	hypothetical protein MGC33215

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
37	9.20E-05	78.565	711.83	204679_at	NM_002245	Hs.376874	KCNK1	potassium channel, subfamily K, member 1
38	0.000101	11.992	180.485	242463_x_at	AI620827	Hs.381287	ZNF28	zinc finger protein 28 (KOX 24)
39	0.000112	410.761	1880.262	31846_at	AW003733	Hs.15114	ARHD	ras homolog gene family, member D
40	0.000113	10.706	111.369	213942_at	AL134303	Hs.56186	EGFL3	EGF-like-domain, multiple 3
41	0.000121	13.812	157.434	214154_s_at	AA888057	Hs.25051	PKP2	plakophilin 2
42	0.000128	27.654	107.156	230076_at	BF063164	Hs.404323	FLJ10156	hypothetical protein FLJ10156
43	0.000133	359.476	696.639	218065_s_at	NM_020644	Hs.389439	C11orf15	chromosome 11 open reading frame 15
44	0.000133	44.485	234.044	227432_s_at	AI215106	Hs.438669	INSR	insulin receptor
45	0.000139	232.897	643.075	209110_s_at	AL050259	Hs.170160	RAB2L	RAB2, member RAS oncogene family-like
46	0.000145	117.65	735.151	212611_at	AV728526	Hs.62264	MPEG1	macrophage expressed gene 1
47	0.000147	39.045	94.608	238953_at	AA993833	Hs.318639		Homo sapiens, clone IMAGE:5206119, mRNA
48	0.000148	35.389	220.792	229223_at	AI038402	Hs.172674	NFATC3	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3
49	0.000153	2.599	51.525	238451_at	BF693302	Hs.350684	FLJ32798	hypothetical protein FLJ32798
50	0.000153	648.584	1252.015	214812_s_at	D80006	Hs.322903	DIP2	disco-interacting protein 2 (Drosophila) homolog
51	0.000157	11.07	61.104	215302_at	AU150691	Hs.153321	LOC257152	hypothetical protein LOC257152
52	0.000161	345.421	819.852	201059_at	NM_005231	Hs.301348	EMS1	ems1 sequence (mammary tumor and squamous cell carcinoma-associated (p80/85 src substrate)
53	0.000166	138.518	491.717	38766_at	AB002307	Hs.136227	SRCAP	Snf2-related CBP activator protein

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
54	0.000169	16.201	264.568	204542_at	NM_006456	Hs.288215	STHM	sialyltransferase
55	0.000171	115.084	294.21	205457_at	NM_024294	Hs.300691	MGC4614	hypothetical protein MGC4614
56	0.000199	8.032	82.637	206604_at	NM_004561	Hs.97905	OVOL1	ovo-like 1(Drosophila)
57	0.000203	106.034	246.146	235093_at	BE564430	Hs.161377		Homo sapiens transcribed sequence with moderate similarity to protein ref:NP_062553.1 (H.sapiens) hypothetical protein FLJ11267 [Homo sapiens]
58	0.000205	263.992	913.944	218284_at	NM_015400	Hs.99843	DKFZP586N	DKFZP586N0721 protein
59	0.000216	161.549	330.118	222732_at	BF514859	Hs.413493	TRIM39	tripartite motif-containing 39
60	0.000218	302.326	1014.514	179_at	U38980	Hs.420556	PMS2L5	postmeiotic segregation increased 2-like 5
61	0.00022	31.773	72.21	223426_s_at	AF153418.1	Hs.207134	EPB41L4B	erythrocyte membrane protein band 4.1 like 4B
62	0.000222	6.76	89.378	243276_at	AA557247	Hs.293637	FLJ36525	hypothetical protein LOC259173
63	0.000224	273.692	1122.075	201349_at	NM_004252	Hs.396783	SLC9A3R1	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1
64	0.00023	3.953	118.909	220318_at	NM_017957	Hs.165904	EPN3	epsin 3
65	0.000238	52.217	126.429	217124_at	AL136792	Hs.446063	KIAA1023	KIAA1023 protein
66	0.00024	379.179	1736.007	203038_at	NM_002844	Hs.354262	PTPRK	protein tyrosine phosphatase, receptor type, K
67	0.00024	3.621	267.872	219850_s_at	NM_012153	Hs.200228	EHF	ets homologous factor
68	0.000241	44.161	202.267	221756_at	AL540260	Hs.26670	MGC17330	HGFL gene
69	0.000243	487.285	987.924	208066_s_at	NM_001514	Hs.258561	GTF2B	general transcription factor IIB
70	0.00025	665.507	1540.641	217837_s_at	NM_016079	Hs.147159	NEDF	neuroendocrine differentiation factor
71	0.000268	66.072	428.367	215243_s_at	AF099730			

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
72	0.000279	173.498	951.179	226789_at	W84421	Hs.446408	na	similar to embigin
73	0.000279	4.765	61.859	207006_s_at	NM_013301	Hs.82482	HSU79303	protein predicted by clone 23882
74	0.000281	221.947	1582.044	226129_at	AI949095	Hs.67776		Homo sapiens, clone IMAGE:5455669, mRNA, partial cds
75	0.000285	258.567	748.169	224871_at	AK025464.1	Hs.20529	LOC127262	hypothetical protein LOC127262
76	0.000286	10.272	993.851	218186_at	NM_020387	Hs.150826	RAB25	RAB25, member RAS oncogene family
77	0.000303	102.766	189.993	213295_at	AA555096	Hs.386952	CYLD	cylindromatosis (turban tumor syndrome)
78	0.000311	243.622	646.01	218018_at	AW449022	Hs.284491	MGC15873	hypothetical protein MGC15873
79	0.000321	186.933	634.329	219241_x_at	NM_017857	Hs.29173	SSH-3	slingshot 3
80	0.000322	54.048	594.387	219395_at	NM_024939	Hs.371804	FLJ21918	hypothetical protein FLJ21918
81	0.000324	7.149	121.205	229901_at	AI056483	Hs.27788	FLJ32104	hypothetical protein FLJ32104
82	0.00033	20.63	133.529	219916_s_at	NM_025236	Hs.121178	RNF39	ring finger protein 39
83	0.000332	18.525	42.258	220270_at	NM_019038	Hs.97464	TDRD4	tudor domain containing 4
84	0.00034	15.1	42.622	210937_s_at	U35632	Hs.32938	IPF1	insulin promoter factor 1, homeodomain transcription factor
85	0.000341	643.43	1624.811	202117_at	BG468434	Hs.410092	KIAA0652	KIAA0652 gene product
86	0.000344	79.353	223.403	244561_at	BG289443	Hs.91389	na	LOC347339
87	0.000344	39.476	526.851	227134_at	AI341537	Hs.25895	JFC1	NADPH oxidase-related, C2 domain-containing protein
88	0.000344	2.006	26.033	229372_at	AW299924	Hs.37477	na	similar to RIKEN cDNA 0610012C01
89	0.000345	12.304	238.606	203918_at	NM_002587	Hs.79769	PCDH1	protocadherin 1 (cadherin-like 1)

p rank	Parametric p-valu	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
90	0.000355	12.265	725.687	238689_at	BG426455	Hs.256897	GPR110	G protein-coupled receptor 110
91	0.000367	437.993	736.854	201933_at	NM_002768	Hs.183138	PCOLN3	procollagen (type III) N-endopeptidase
92	0.000385	79.699	178.621	204613_at	NM_002661	Hs.271620	PLCG2	phospholipase C, gamma 2 (phosphatidylinositol-specific)
93	0.000389	20.433	147.413	204378_at	NM_003657	Hs.400556	BCAS1	breast carcinoma amplified sequence 1
94	0.00039	41.505	178.488	219503_s_at	NM_018306	Hs.263876	FLJ11036	hypothetical protein FLJ11036
95	0.000397	59.696	340.286	210026_s_at	AY028896	Hs.57973	CARD10	caspase recruitment domain family, member 10
96	0.000403	30.778	68.553	217325_at	X05421			
97	0.000404	51.677	200.038	218523_at	NM_022126	Hs.20950	LHPP	phospholysine phosphohistidine inorganic pyrophosphate phosphatase
98	0.000407	151.805	758.287	225612_s_at	BE672260	Hs.257222	B3GNT5	UDP-GlcNAc:betaGal beta-1,3-N acetylglucosaminyltransferase 5
99	0.000424	219.594	438.115	219238_at	NM_017837	Hs.259605	FLJ20477	hypothetical protein FLJ20477
100	0.000446	4.216	190.025	212338_at	AA621962	Hs.39871	MYO1D	myosin ID
101	0.000456	7.255	455.005	219936_s_at	NM_023915	Hs.58561	GPR87	G protein-coupled receptor 87
102	0.000468	79.122	392.202	227533_at	AA732944	Hs.446665		Homo sapiens transcribed sequences
103	0.000475	96.173	199.291	218743_at	NM_024591	Hs.22897	FLJ11749	hypothetical protein FLJ11749
104	0.000482	7.925	97.382	235050_at	AI742872	Hs.26691	SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12
105	0.000482	6.401	184.857	241455_at	AW135306	Hs.444277		Homo sapiens transcribed sequences
106	0.000503	12.231	220.119	202994_s_at	Z95331	Hs.445240	FBLN1	fibulin 1

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
107	0.000512	383.478	901.639	212622_at	N64760	Hs.174905	KIAA0033	KIAA0033 protein
108	0.000519	98.014	320.235	218815_s_at	NM_018022	Hs.30925	FLJ10199	hypothetical protein FLJ10199
109	0.000527	43.113	283.207	203408_s_at	NM_002971	Hs.416026	SATB1	special AT-rich sequence binding protein 1 (binds to nuclear matrix/scaffold-associating DNA's)
110	0.000535	10.341	381.817	206504_at	NM_000782	Hs.89663	CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1
111	0.000536	23.595	151.332	222879_s_at	AF158185.1	Hs.155573	POLH	polymerase (DNA directed), eta
112	0.000536	50.745	148.913	214763_at	AK023937	Hs.234786	THEA	thioesterase, adipose associated
113	0.000551	11.168	364.378	226535_at	AK026736.1	Hs.57664	ITGB6	integrin, beta 6
114	0.00057	85.569	1718.174	203407_at	NM_002705	Hs.192233	PPL	periplakin
115	0.00058	15.309	123.404	207655_s_at	NM_013314	Hs.167746	BLNK	B-cell linker
116	0.000582	57.597	565.098	223103_at	AF151810.1	Hs.300446	STARD10	START domain containing 10
117	0.000601	12.985	121.911	238725_at	AW392551	Hs.180559		Homo sapiens transcribed sequence with weak similarity to protein ref:NP_055301.1 (H.sapiens) neuronal thread protein [Homo sapiens]
118	0.000624	51.901	329.095	219429_at	NM_024306	Hs.132380	FAXDC1	fatty acid hydroxylase domain containing 1
119	0.000699	54.096	983.946	218677_at	NM_020672	Hs.288998	S100A14	S100 calcium binding protein A14
120	7.00E-04	256.84	938.775	200752_s_at	NM_005186	Hs.356181	CAPN1	calpain 1, (mu/I) large subunit
121	0.000702	69.408	193.722	209454_s_at	AF142482	Hs.203846	TEAD3	TEA domain family member 3
122	0.000759	9.804	143.91	219233_s_at	NM_018530	Hs.306777	PRO2521	hypothetical protein PRO2521

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
123	0.00076	3.006	247.885	203638_s_at	NM_022969	Hs.404081	FGFR2	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, Crouzon syndrome, Pfeiffer syndrome, Jackson-Weiss syndrome)
124	0.000918	7.971	201.676	235651_at	AV741130	Hs.173704		Homo sapiens transcribed sequence with moderate similarity to protein ref:NP_060265.1 (H.sapiens) hypothetical protein FLJ20378 [Homo sapiens]
125	0.000994	80.562	4.126	230261_at	AA552969	Hs.308628	SIAT8D	sialyltransferase 8D (alpha-2, 8-polysialyltransferase)
126	0.000956	256.802	63.684	222565_s_at	BF978541	Hs.434387	PRKCN	protein kinase C, nu
127	0.00095	39.569	3.235	205591_at	NM_006334	Hs.74376	OLFM1	olfactomedin 1
128	0.000933	1996.522	1059.755	211275_s_at	AF087942	Hs.174071	GYG	glycogenin
129	0.000892	137.586	29.447	220370_s_at	NM_025090	Hs.11387	KIAA1453	KIAA1453 protein
130	0.000885	165.637	26.581	218976_at	NM_021800	Hs.260720	JDP1	J domain containing protein 1
131	0.000828	855.192	489.029	203025_at	NM_003491	Hs.433291	ARD1	ARD1 homolog, N-acetyltransferase (S. cerevisiae)
132	0.000818	487.076	136.446	221489_s_at	W48843	Hs.406507	SPRY4	sprouty homolog 4 (Drosophila)
133	0.000787	1678.035	788.047	200842_s_at	AI475965	Hs.171292	EPRS	glutamyl-prolyl-tRNA synthetase
134	0.000776	51.574	10.419	239790_s_at	AW292017	Hs.211973	RRP4	homolog of Yeast RRP4 (ribosomal RNA processing 4), 3' 5'-exoribonuclease

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
135	0.000766	324.81	155.056	203358_s_at	NM_004456	Hs.444082	EZH2	enhancer of zeste homolog 2 (Drosophila)
136	0.000765	189.807	18.79	219093_at	NM_017933	Hs.424598	FLJ20701	hypothetical protein FLJ20701
137	0.000761	339.385	34.38	212736_at	BE299456	Hs.148258	BC008967	hypothetical gene BC008967
138	0.000759	495.844	85.546	219270_at	NM_024111	Hs.155569	MGC4504	hypothetical protein MGC4504
139	0.000721	407.011	204.561	223266_at	AB038950.1	Hs.259230	ALS2CR2	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2
140	0.000717	160.769	65.588	212684_at	AI752257	Hs.435302	ZNF3	zinc finger protein 3 (A8-51)
141	0.000706	1251.634	560.836	218715_at	NM_018428	Hs.211828	HCA66	hepatocellular carcinoma-associated antigen 66
142	0.000699	290.065	130.744	222239_s_at	AL117626	Hs.396557	DDX26	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26
143	0.000685	237.246	142.75	210053_at	AW138827	Hs.171625	MGC14697	upregulated during skeletal muscle growth 5
144	0.000674	194.051	18.439	202340_x_at	NM_002135	Hs.1119	NR4A1	nuclear receptor subfamily 4, group A, member 1
145	0.000662	64.592	14.968	230327_at	AI203673	Hs.225948	CCL27	chemokine (C-C motif) ligand 27
146	0.00066	398.499	184.454	204853_at	NM_006190	Hs.167937	ORC2L	origin recognition complex, subunit 2-like (yeast)
147	0.000655	34.116	4.283	244091_at	AI560305	Hs.199852		Homo sapiens transcribed sequences
148	0.000635	46.775	5.893	235567_at	AA034012	Hs.37648		Homo sapiens cDNA FLJ31407 fis, clone NT2NE2000137.
149	0.000634	162.772	80.338	239033_at	AI640482	Hs.283986		Homo sapiens mRNA; cDNA DKFZp666M079 (from clone DKFZp666M079)

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
150	0.000632	3567.711	1214.664	215111_s_at	AK027071	Hs.114360	TSC22	transforming growth factor beta-stimulated protein TSC-22
151	0.000612	157.791	30.712	221168_at	NM_021620	Hs.287386	PRDM13	PR domain containing 13
152	0.000585	7037.683	4778.217	200705_s_at	NM_001959	Hs.421608	EEF1B2	eukaryotic translation elongation factor 1 beta 2
153	0.000561	175.69	4.49	204944_at	NM_002841	Hs.89627	PTPRG	protein tyrosine phosphatase, receptor type, G
154	0.000544	105.146	29.932	205993_s_at	NM_005994	Hs.168357	TBX2	T-box 2
155	0.000526	133.741	89.35	239077_at	W81648	Hs.180758	GALNACT-2	chondroitin sulfate GalNAcT-2
156	0.000498	764.21	355.399	215136_s_at	AL050353	Hs.274170	OIP2	Opa-interacting protein 2
157	0.000486	197.4	36.042	228962_at	BF507941	Hs.28482	PDE4D	phosphodiesterase 4D, cAMP-specific (phosphodiesterase E3 dunce homolog, Drosophila)
158	0.000475	107.437	9.284	211323_s_at	L38019	Hs.149900	ITPR1	inositol 1,4,5-triphosphate receptor, type 1
159	0.00046	65.742	3.851	231721_at	AF356518.1	Hs.419149	JAM3	junctional adhesion molecule 3
160	0.000447	2328.925	889.565	201490_s_at	NM_005729	Hs.381072	PPIF	peptidylprolyl isomerase F (cyclophilin F)
161	0.000445	5719.836	3769.693	201568_at	NM_014402	Hs.146602	QP-C	low molecular mass ubiquinone-binding protein (9.5kD)
162	0.000423	575.737	27.962	207325_x_at	NM_004988	Hs.72879	MAGEA1	melanoma antigen, family A, 1 (directs expression of antigen MZ2-E)

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
163	0.000398	53.27	8.647	243680_at	AI650285	Hs.287299		Homo sapiens transcribed sequence with weak similarity to protein ref:NP_060312.1 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]
164	0.000392	631.178	48.094	214240_at	AL556409	Hs.278959	GAL	galanin
165	0.000365	1595.002	587.109	209017_s_at	U02389	Hs.350265	PRSS15	protease, serine, 15
166	0.000346	491.002	229.889	222894_x_at	AI640582	Hs.420282	C20orf7	chromosome 20 open reading frame 7
167	0.000338	955.064	384.915	204957_at	NM_002553	Hs.153138	ORC5L	origin recognition complex, subunit 5-like (yeast)
168	0.000336	745.787	32.893	204105_s_at	NM_005010	Hs.7912	NRCAM	neuronal cell adhesion molecule
169	0.000315	2129.962	1134.195	208968_s_at	BC002568	Hs.4900	LOC57019	hypothetical protein LOC57019
170	0.000285	252.557	18.687	204513_s_at	NM_014800	Hs.444695	ELMO1	engulfment and cell motility 1 (ced-12 homolog, C. elegans)
171	0.000278	169.662	72.892	229063_s_at	AI912238	Hs.245342	FLJ14642	hypothetical protein FLJ14642
172	0.00025	650.067	179.374	202146_at	AA747426	Hs.7879	IFRD1	interferon-related developmental regulator 1
173	0.000245	517.157	175.417	224896_s_at	AI888594	Hs.358997	TTL	tubulin-tyrosine ligase
174	0.000232	68.258	22.873	232080_at	AL390186.1	Hs.210381	NEDL2	NEDD4-related E3 ubiquitin ligase NEDL2
175	0.000216	986.619	448.192	209482_at	BC001430	Hs.416994	RPP20	POP7 (processing of precursor, S. cerevisiae) homolog
176	0.000212	521.649	294.879	228095_at	AA608749	Hs.409117	PHF14	PHD finger protein 14
177	0.000212	23.173	2.916	241729_at	AW173080	Hs.303172	MGC20785	hypothetical protein MGC20785

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
178	0.000201	373.198	176.83	225311_at	AA081349	Hs.410396	IVD	isovaleryl Coenzyme A dehydrogenase
179	0.00019	940.929	530.053	210719_s_at	BC002552	Hs.406534	HMG20B	high-mobility group 20B
180	0.000181	641.453	342.18	221692_s_at	AB049652	Hs.238808	MRPL34	mitochondrial ribosomal protein L34
181	0.000181	187.58	11.76	219855_at	NM_018159	Hs.200016	NUDT11	nudix (nucleoside diphosphate linked moiety X)-type motif 11
182	0.000173	923.269	387.005	226434_at	BF000655	Hs.413359	MGC22793	hypothetical protein MGC22793
183	0.000163	255.268	12.365	236741_at	AW299463	Hs.208067		Homo sapiens transcribed sequences
184	0.000156	3353.919	2041.548	209248_at	AL136713	Hs.352656	GHITM	growth hormone inducible transmembrane protein
185	0.00011	911.17	8.581	205174_s_at	NM_012413	Hs.79033	QPCT	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)
186	9.20E-05	193.518	16.424	209621_s_at	AF002280	Hs.71719	ALP	alpha-actinin-2-associated LIM protein
187	8.80E-05	2113.683	841.359	201682_at	NM_004279	Hs.184211	PMPCB	peptidase (mitochondrial processing) beta
188	8.80E-05	6125.903	118.574	216442_x_at	AK026737	Hs.287820	na	similar to hypothetical protein FLJ20958
189	8.50E-05	374.31	2.782	206710_s_at	NM_012307	Hs.103839	EPB41L3	erythrocyte membrane protein band 4.1-like 3
190	5.40E-05	6459.858	2394.517	224841_x_at	BF316352	Hs.289721	LOC348531	hypothetical protein LOC348531
191	4.50E-05	163.384	83.008	231297_at	AI479899	Hs.289848	DOT1L	histone methyltransferase DOT1L
192	4.30E-05	1875.625	669.002	209825_s_at	BC002906	Hs.458360	UMPK	uridine monophosphate kinase

p rank	Parametric p-value	Geom mean of intensities in class 1: resistant	Geom mean of intensities in class 2: sensitive	Probe set	GB acc	UG cluster	Gene symbol	Description
193	4.20E-05	2370.118	1062.626	222997_s_at	BC004566.1	Hs.405880	MRPS21	mitochondrial ribosomal protein S21
194	1.30E-05	3735.302	2156.379	200652_at	NM_003145	Hs.74564	SSR2	signal sequence receptor, beta (translocon-associated protein beta)
195	1.10E-05	3989.574	73.169	212464_s_at	X02761	Hs.418138	FN1	fibronectin 1
196	9.00E-06	335.981	30.024	212624_s_at	BF339445	Hs.380138	CHN1	chimerin (chimaerin) 1
197	9.00E-06	57.278	6.976	215517_at	AL049925	Hs.256587	PYGO1	pygopus 1
198	7.00E-06	310.353	116.392	212471_at	BE503381	Hs.134792	KIAA0241	KIAA0241 protein